MEDICAL IMAGING STAKEHOLDERS CALL FOR ACTION:
HARMONIZATION OF IMAGING REVIEW CHARTERS AND INTEGRATION OF IMAGING IN THERAPEUTIC DEVELOPMENT:
PHARMACEUTICAL INDUSTRY, CRO, FDA AND ALLIED WORKING GROUPS COLLABORATE FOR REGULATORY GUIDANCE

MEDICAL IMAGING CONFERENCE PROCEEDINGS

DECEMBER 4, 2007

COLLABORATING ORGANIZATIONS:

Prepared by: Medical Imaging Program Committee

(MIWG) Representatives Pharmaceutical, CRO, Academia Industries, Regulatory and Allied Working Groups

Drug Information Association (DIA):

Constance Burnett, Meeting Organizer and Program

8:07 a.m. through 4:42 p.m.

October 16-17, 2007

Marriott Conference Center

University of Maryland

Adelphi, Maryland

Table of Contents

P R O C E E D I N G S

Welcome and Opening Remarks

DR. MOZLEY: Good morning, everyone. It is good to have you back. We are looking forward to another productive day.We have a lot of ground to cover. I think it is possible to cover it quickly and effectively.
So, without further ado, we would like to please invite the Subgroup No. 3 on the site interface to take over the show and lead us through the day.
Thank you.

Session 3

Panel Discussion Leaders

Dr. David Clunie, Chief Technical Officer, Radpharm
Mr. Stefan Bauman, Imaging Infrastructure Manager
Dr. Ed Ashton, Chief Scientific Officer, VirtualScopics
Dr. Aldo Badano, Director, Imaging Physics Laboratory, FDA

Key Points to Consider for Core Laboratories

DR. CLUNIE: Good morning, everybody.
Group No. 3 focused on the site core lab interface, and we have the next hour and a half to go through our material. So we will follow the same tactic as was used yesterday. We will essentially just step through our major topics until we either run out of time or you guys get sick of us, and then at 9:45, we will stop, break, and let the statisticians take over for the next session and see how much we have covered in terms of what needs to be addressed after lunch, if anything.

I would like to welcome to the table here, Ed Ashton and Stefan Baumann from Novartis. Ed is from VirtualScopics, and Aldo Badano, who represents the Imaging Physics group at the FDA.
In our early discussions after the first face-to-face meeting that we had earlier in the year, we identified a number of areas being relatively high priority of concern.

Deidentification came up as a concern because, as we do more of these multinational trials, we run across various different directives and legislation in different countries that is making it increasingly difficult to deal with the material we get from the sites.

We talked about the image transfer. We talked about whether or not one could use lossy compression. Digitization of film is a topic that we wanted to address, albeit briefly, a matter of retention, whether there are standards for how long material should be retained, what material should be retained at what stage after the trial

has been completed and the drug has been approved, the matter of reuse of information gathered during the course of a clinical trial, reused for other purposes, site initiation and monitoring, and validation of software. Some of these, we went into a lot more detail than others, and that will become apparent as we go through the morning.

We set out to decide what we thought was in scope and what we thought was out of scope, and in a similar manner to what you saw from the teams yesterday, we decided that we would focus on registrational trials, we would focus on medical imaging, we would focus on studies in humans only, of course. In vivo imaging would be our primary focus.

We would include image-like data, other kinds of bulk data that are acquired from imaging acquisition devices such as MR spectroscopy. We only wanted to focus on mature technologies rather than exotic things that might be going through device approval, such as optical tomography or whatever, and we also decided that if phantoms are to be acquired and phantom images need to be transferred, essentially, that those would be in scope because they are essentially a site interface issue.

What we concluded would be out of scope would be anything to do with early drug development, in vitro imaging, experimental technologies, transfer of modality protocols, and by that I mean transfer of the instructions used to make the machine execute a particular type of imaging as opposed to the images themselves, and there are some efforts to standardize that, but we decided that was out of scope at this time. It is too cutting-edge, shall we say.

Phantom requirements, we decided would be out of scope, although, of course, we can't resist mentioning phantoms, and you will see that later.
The matter of submission to regulatory agencies, in what form, if any, should the images and their accompanying data be submitted, we decided that was out of scope because we are only interested in the site to CRO or site to sponsor interfaces at this point.

Our methodology is the same, trying to achieve consensus between the various players who identified themselves as stakeholders, identify the areas of concern, prioritize these, and we have been fairly selective about the amount of depth we put into this based on what we thought would be the return on investment, as well as the likelihood of consensus, and essentially where we had no consensus or we thought it was going to be controversial, we excluded it.

If necessary, propose standards or recommendations or guidelines or good practice or whatever we want to call it, and I guess it has yet to be determined what we were going to actually call the output of this process prior to the FDA guidances, and then hopefully, at some stage, adopt, promulgate, and promote whatever it is we decide upon.

We had quite a list of contributors. You have seen these in your other material, but presumably, you will be able to download the slides later. We were very pleased with the contribution that various people made on a series of telephone conferences.

So, in terms of getting to the details, I am going to give a very broad overview of each of the key topics and essentially quote from the document how we set up the parameters for each of these priority areas, and then each of the subsequent speakers will go into more depths in each of these in terms of our recommendation and references to the document.

Deidentification is a constant and irritating problem due to, essentially, gratuitous variations between different sites' policies, different sponsors' policies, different CRO policies, and the variation in expectations and the quality of the work and the lack of audit trails is causing us a problem. So we will cover that in probably more detail than any other topic.

Transfer from the sites, although named as a primary source of pain by most people when they first think of this, is very difficult to standardize. It is debatable whether this is in the precompetitive space or whether this is a source of competitive advantages for the companies that specialize in the logistics, and also it is something that may just solve itself in time. We will come back to that later, but just to highlight the fact of standardizing the manner of transfer was not targeted as an area that we thought we were going to make significant achievements with in this document.

Compression is almost a spurious topic in a way, but one which is unavoidable because sites and others continue to occasionally insist on using it. The bottom line at the moment is that lossy image compression translates into protocol violations, and we tried to look at some aspects of that to see if we could rationalize the discussion on this, and we will discuss that in more detail later.

Digitization of film, which everybody regards as undesirable, is inevitably unavoidable under some circumstances, and the question arises can we or do we need to have any quality and performance standards related to this, is there really a void of standards in this area, is quality improvable in this area.
Retention. People confuse the clinical requirements for retention that are imposed by local and State and

Federal regulations with the requirements for clinical trials retention.
So we tried to clarify the distinction and go into some detail as to what should or should not be retained and under what circumstances and for how long, and that will be addressed in a bit more detail.
Site initiation and monitoring is obviously a very important part of managing any imaging clinical trial, are there any areas where we can achieve some degree of consistency here, so that the sites aren't faced with variation between CROs and sponsors, and likewise we can achieve consensus in terms of how we initiate multiple sites, to what extent is monitoring necessary, and that will be addressed specifically by one of the speakers.
Validation of software is obviously a hot topic and the thing that ostensibly distinguishes commercial trials from academic trials, and the dreaded 21 CFR Part 11 raises its head, but remember that sites have no understanding of 21 CFR Part 11, and if the site is not using approved medical devices, then this become potential issue. We will discuss this briefly.

We were instructed to produce a lexicon, and this turned out to be not a bad idea, although we have focused primarily in the lexicon at this point on exotic terms related to the matter of deidentification. There are probably a number of other terms we could factor out and produce a better lexicon.

There may be some slight overlap with other groups' work on lexicons, which we can expand upon and harmonize on, and still to do, probably, we have to consult the literature and other documents to make sure that our definition of terms for these particular terms listed matches what is commonly understood in the literature, and by that, I mean also outside the clinical trials industry, sort of the general health care IT industry has its notions of what these terms mean, quite apart from the imaging world. So we need to make sure that we are consistent in that respect.

It is also very difficult to come up with terminology that is consistent between different countries. The terminology that is used in the U.S. differs quite significantly from the terminology that is used in the European privacy directives and the various different member states' enactment of that. This is actually quite challenging to address, and just because we define it one way doesn't mean that those definitions will be generally applicable. So I think this is an area where we need to do a little bit more work and clean up the document beyond what you will see on page 23.

I have tried to tag some of these slides with references, in case you are following along in the document or want to review the material in more depth a bit later.

So, at this point, Stefan will address the deidentification issues. Ed Ashton will address site initiation monitoring after that. Stefan will also talk about retention, and then I will finish off with addressing some of the other topics.

At any point, we would welcome your comments. So, after each topic section, we will raise specific discussion points that address things that the FDA and their comments or others have raised or, alternatively, things that we have saw preparing these slides or that those of you who have mentioned issues to us in the corridor have wanted addressed.

Before I proceed, though, I want to make sure if there are any issues arising out of yesterday's discussion that don't really have a specific place on the agenda today, that we give people an opportunity to discuss those, and I seem to recall that Lou Marzella wanted to make some comments related to yesterday's work.

DR. MARZELLA: Thank you.
I just wanted to reiterate the idea that we are still working on defining the pieces that we need to do a complete review of an imaging submission. So it would be important for us to understand how your group would sort of think about this, about how to organize and link the various components of an IRC to make sure that we can have access in a sort of seamless fashion as possible to all the various components that we need to do the regulatory review.

DR. CLUNIE: So you are referring specifically to the documents that are available to a review at a particular state. So whoever is reviewing the charter needs the protocol. Whoever is reviewing the statistical analysis plan needs the charter and the protocol, et cetera, et cetera.
One way potentially we could deal with that in the document organization -- and this is certainly outside the scope of our particular subgroup, but the obvious idea that comes to mind is that perhaps in the table of contents section, we should actually lay out in our recommendations for documentation practices, the sequence in which the reviewers need the material and the packet table of contents that they need at each particular point of the review, and maybe that is something everybody could agree on, on both sides and follow that practice.
Delay of review is something everybody wants to avoid at all costs, obviously. Yes.
Are there any other comments about yesterday's work, or should we just proceed?

MR. BAUMANN: Good morning. I will walk you through the process of the anonymization that we have looked at very closely. We started with identifying the issues that are today hindering a smooth process of anonymization of images.

We have various regulations and various varying privacy expectations in hospitals, IRBs in country-wise, and even on the EU versus the U.S., where in EU laws, we have even much stricter laws about data export. If we speak about sensitive personal elements, we have privacy guidelines in the EU which puts down very strict rules on what you can do with the data and how you have to prepare such activity.

Lack of effective standards, we have DICOM standards that give some guidelines about how to anonymize DICOM images, but we would have to put effort to extend that to really match the full needs of the clinical trials. We have relatively uncontrolled process at the site where normally it is the site obligation to anonymize the images before they send it out, but there we have a lot of variability, and some sites will do some deidentification process. Some other sites will not do anything, and it is really difficult to control which part of a site's anonymization process is really positive versus which part of the process is taking out too much information that we subsequently still need for the evaluation. The quality of work is basically very closely connected to that. It is very valuable.

We have put down a list of things that really should be anonymize when we are speaking of anonymization, what are the items that we do not want to carry on within the images. So the images, as you all know, they have an actual bitmap, and then they have a lot of header information. In both of them, we find information that can possibly be identified in the identity of someone.

So, as a first item, I have deidentifiers which are called sensitive and sensitive personal information in EU laws. Site-related information, equipment-related information, physician-related information, these are all things that can go in a DICOM image and should be put out before it sends to the reviewer.
We have a second category of potential identifiers that are not normally thought to be this close in a patient's identity, but in one or the other case, quite often it can happen that some identifying information happens to be in there, just because someone decided to put it there.

A third category which is very closely related is proprietary and unknown information for all the manufacturers, proprietary tax which are transferred together with the imager. It is a high risk of identifying information to be present in there.

The image unique identifiers is another topic in itself. Every image is accompanied by a globally unique identifier which is such a very nice thing to put on the back side, on the flip side of it. An image, if you look at the unique identifier, can be traced back to exactly where it was provided and at what event it was provided and obviously also identifying the patient. So we have to find mechanisms to protect those.

We have to make sure that no information is on the bitmap of the image. So it happens quite a lot, and a technician makes an image that they have some information about a patient at the display and when they save it to a digital format, they will save this information on the bitmap together with the image.

On the other hand, we have some variables that are critical to look at, but we have to maintain those in order to

do a full valid analysis, and first of all, these are exam dates. There is the HIPAA privacy rule which doesn't permit dates to be present if they can be linked to an event, but in this case, it is really difficult to get rid of them. It is difficult to round them up or to have a certain offset date which we can add to that because that can add very much complexity to imaging trials, also to trials where we have clinical endpoints and where we need to compare the date of acquisition. So we need to keep those dates in.
For the unknown information tags or for the known tags where we cannot be sure whether they contain the identifying information, it is best to make sure that we have some pattern checks. For example, if there is a date, we can have a pattern check to make sure, and on the incoming image, a date is really a date and doesn't contain letters.

We have as a third category identifiers for specific study types. We talk about that later again. For SUV, we need to collect the sex and the weight of the patient. So this is surely something which is an exception in this case and which needs to be collected.

The last category that we noted down here are reconciliation identifiers. It is common practice for image evaluations in longitudinal studies to compare a couple of additional identifiers next to the anonymized subject Ids in order to make sure that no confusion of images has happened, since that we are having a clean dataset.

So this is very critical information which we need to find out, and we haven't yet in full detail specified that, how many identifiers can we keep for reconciliation and which ones specifically can we keep. EU laws here again are very critical where it is not simply possible to capture more than the really essential minimum
Some notes to anonymization, the informed consent needs to be very specific. Depending on the anonymization process, it does need to contain personal information, sensitive personal information that should be transferred, like if that was to be smoking status or ethnicity or something like that. This needs to be on the IC.

Also, for data export, if data is to be exported outside of the EU and it is acquired within the EU or it is read outside of the EU or outside of safe harbor states, then this has contractual obligations, and it also needs to be stated on the IC.

The EU privacy rule also specifically says that if you have something that is already collected in clinical records, then you should not collect that. Again, we have to go into more detail to which extent that is a lawyer's interpretation and to which extent we can take that out, like this directly from the source, but in general, we recommend if there is some variable like a smoking status, which you could transport an image, but you have collected it in the clinical data, then we recommend to avoid to collect it again in the image data set. That also has to do with variance, obviously. So it is nice to have only one source data, one data source, and it has with it

the EU privacy law.

When removing identifiers specifically with age, date of birth, we can also talk about rounding functions. There are very commonly used also for reconciliation, identifiers like age.

You can remove identifying aspects of this by rounding it, or in some other case, hashing which means transforming the identifier that should be protected into an unidentifiable element, but doing that in a way that it is always reproducible.

As a last point -- and this is an open point which is tabled for future consideration -- three-dimensional reconstruction of faces from high-resolution MR images or other types of images are a risk which is currently quantified and is at this moment out of scope.

The way forward, in order to dig into the detail and come up with exactly what should be out of those more than 4,000, I think, DICOM attributes that potentially contain or do not contain identifying information, in order to exactly define which of those should be protected, which should not, we want to extend the DICOM standard to 3.15 for clinical trials and make sure it is up to date on the very recent status of the potential elements that can go into an image.

The group also has decided to go for an approach where we have profiles. We will try to go for a base profile which is general anonymization, and then add exceptions on the first study type. So, for example, FDG PET, you will have a profile declaring that. In this case, you also need to collect sex and the weight, and this will make it easier for the necessity or collecting such identifiers, and you can relay back to a standard which supports your need for those collections.

The process quality is a segue discussion which will lead into the retention. This is a mixed somehow slide about anonymization and retention.

The image transfers between the site modalities should not affect image consistency, and to give an example of that is if you have PACS which is an image management system at a hospital, if this PACS is used to manage all image transfers between a modality and potential received in a clinical trial, then it is sometimes possible that those PACS systems remove elements that are vital for us to collect for certain evaluations. We need to at least come up with a rule that states that image transfers should always be unaffected by any kind of management inside a site.

The CRO has a role to monitor the anonymization process quality-wise and is still responsible for the anonymization, and while it is our goal to bring as much as possible, regulations to the sites, that unifies this process of sites, anonymization. The CRO still has an important role to monitor this and work around for issues where it is not followed properly.

We have discussed also local audit trail for anonymization. The audit trail in this case is a little critical because if you remove items in an image and you want to maintain an audit trail of this activity, you can obviously not transfer this to the sponsor because it can contain the nature of those elements.
A nice discussion that we had about that was for lacking electronic management at the site for those audit trails, we could actually have a process which produces a paper audit trail which is stored together with the source data.

Finally, another point which is a segue into retention is the site archive of the source data. The site is obliged to keep an archive of the source data, and we will have to talk about source data and the nature of source data. Actually, before I continue to retention, I want to ask if there is any comment or any questions about the anonymization.

MR. BAUMANN: Okay. Then I will continue with the retention.
Good practice. One thing that looks trivial, but it is very important to enable a better process in image management and analysis. The images should be always stored in best-of-breed standard which is most of the time DICOM standard. There can be instances where it provides too much overhead. It gives too much overhead to convert something, maybe like a visual light image into DICOM form, while it is still possible for some study set-up, it may be too high an overhead, but at least whatever is stored, it should always be in the best-of-breed format, which could be like a .jpeg or a .tif as long as it is not lossy compression, which we come back to that later, or we will have to look at that in detail later.

This is a very straightforward point that the format should always be DICOM or best of breed because, if you go back in 10 years and want to retrieve the image, it will be impossible if it is not this way.
Then we have two types of audit trails that we want to propose. We have the deidentification audit trail that we will talk about later, and we should have an audit trail for the transfer to the CRO.

So here are a couple of regulations that speak about retention obligations. You have clinical regulations for retaining clinical routine care images versus trial requirements for the site. We have GCP and CFR regulations which say that the image data as essential source data has to be retained until after two years after approval.

You have SOPs that further sponsor and site and CRO and further detail, the use of those, the timelines of the retention. The most important question about that is what to retain.

In order to address this point, I want to have a segue into what is source data. For images, it is not completely trivial to discuss what is the source data. Oftentimes it is more easy if we have paper documents and we convert them into electronic format. Then it is easy to have an entry point of this conversion, have that be signed with possibly an electronic signature, and then maintain an audit trail from this moment in imaging where it starts. With electronic files, it is more difficult to determine.

The question is, is an image source data as it comes from the clinical care environment, or is it only source data after the point of anonymization when it enters the world of clinical trials. So, to remind you again, the image stays the same before or after the anonymization. It is just a descriptive header information that locates certain patients to the image, which will change before and after deidentification.
In terms of retention requirements, it is also clear that the site has the clinical burden to retain and as well as a GCP burden to retain, and the CRO normally has higher experience in how to retain data and can do that in a more controlled process, but can only retain what is sent to them.

So if we look at that from a more graphical perspective, we can start with an image generated at the scanner, and then we can go to our anonymized image which is at the bottom at the right. So the writing convention is that we have an image. This is this black thing, and on top of it, we have a symbol that describes the header information which comes together with the image.
You see here that depending on the site, it will always be a different process how we get from an image to an anonymized image. Sometimes we can go directly from the scanner to an anonymized image and then the CRO.

This is then the top middle. This is an image as it is directly anonymized from a site, and we go then down to the final anonymized image because oftentimes the site will do an anonymization procedure which may not be enough or not be coherent with the sponsor's expectations.

Another way that can happen if the site is very well compliant is that from the scanner, we go directly to the anonymized image which remains the best anonymized image which will not be changed anymore, and the third process which is possible is that the image goes first into the site PACS which is the routine archive for all the clinical care images, and only then in a second step, the header information will be changed for anonymization.
Then we have even this relation to film data, which also are sometimes subject to archive.
So the question here is where is the source data. The question is to half motivated by the different variances in

the process and to the other half is motivated by figuring out whether the identifier or the non-identified image is the key source data image.

In addition to this, I will walk you though the archive discussion points that came up in general. The core lab archiving responsibilities, it will, of course, depend heavily on the previous discussion about the source data and about what is sent from the site to the CRO and how well controlled the process this has been.
Physical CD-ROM archive in addition to site archive, so if the site sends a copy of their CD-ROM to the core lab, what is the core lab's obligation to retain this very physical CD as opposed to just have an online archive of readable image versions which they read from the CDs and even potentially have to do some conversions in order to make it fully readable and archivable.

Normally, you would expect only one retention copy per image version, which means that if there is some subsequent anonymization process, you come up with a new image, so to say, where you have to store either the delta of the conversion or the new image. In an idea process, next to the site source data archive, you would not have a need for an additional source data copy. That is in an ideal world at the CRO.

What is the process for films? Where the original films sometimes have to be returned to the site, and the core lab may have a digitized copy for their archive.

What is the form of the archive? Are there any rules on media or device? It does maybe not make sense to store something for 10 years on a DAT tape where you would expect that after 10 years, it is more difficult to for that DAT tape than it is for other types of media, so should we come up with recommended media or device.
Potential technical issues with PACS, that was mentioned previously. How can we make a best effort to ensure that at the site, we do not deliberately lose information by sending the image through a couple of image management systems that may alter the image before it is sent to the CRO for evaluation to the core lab?
The film archive, this is a relatively obvious point. It is not allowed for digitally imprinted images. I skilled lossy compression. So there shouldn't be a case where after the evaluation, another subsequent step of lossy compression is done before archiving.
So that is all about retention on the archive. Is there any questions, any input to this?

DR. ANALOUI: I have a question regarding retention. As you mentioned, there is different requirement for the clinical retention versus clinical trial retention of data, that two year, either post approval or discontinuation.
Do you see that potentially that the clinical retention time period is much longer than clinical trial? Do you see any chance that that longer term retention could come into the clinical trial and make that requirement much longer than two years, especially in a scenario that a CRO could have clinical feedback back to the sites that is related to the

patient care?

MR. BAUMANN: I think one aspect, as we interpret this two years, also to note is that this refers to two years after any subsequent activity is done with those images. It could be also a result that comes from reuse of the data to make further proofs about a hypothesis.
Usually, in sponsor SOPs, these two years result really in something along the lines of five to maybe even eight to 15 years of retention period.
Maybe, David, you want to say something about that, also.

DR. CLUNIE: Yes. The retention periods are defined for two different reasons. In the case of GCP and CFR as they pertain to clinical trials, the idea is to be able to reconstruct the trial for some interval after approval.

Whereas in the clinical world, it is to provide either better clinical care or to address medical-legal concerns and limitations on those, and that varies tremendously from jurisdiction to jurisdiction, but seven years is typical in the U.S.

The other thing to bear in mind is that in many countries, there aren't such retention requirements. Also, the retention requirements are sometimes addressed by giving it to the patient. So, in other words, the facility doesn't retain the information. They give it to the patient and say, "It's your problem." So that means the retention period ranges from zero to something quite belonged.

With respect to the question of findings on the image subsequently, I think people seem to agree yesterday that that was a really dangerous track to go down, that at all costs, CROs and sponsors should disclaim any duty, shall we say, to the patient in that respect, and that as a consequence, there should be no expectation of retention for that purpose.

DR. CHARLES: In addition to that, at some point at the site, they are going to destroy the link to whatever code they have used in the image, and you are not going to be able to reestablish that link.

DR. CLUNIE: That is true to the extent that nobody retains the relationship between the patient's true identity and identity on the trial, but I guess sponsors are a bit mixed in respect to that. Some sponsors know who the patients are. It has always been traditional to know who the patients really are, and it is only very recently that some people have tried to distance themselves from patients' true identities in terms of in-house sponsor information, but there is certainly no question that you can't rely on the principal investigator at the site for any

length of time, as far as I can tell, except with respect to their GCP responsibilities which they may not take seriously.

DR. CHARLES: Going on from that, the concept at the site as the ultimate location of the digital source data is sort of a sketchy thing, too, because depending on how they are storing the data, like whatever particular medium of the future, like the 12-inch optical platters, by the time you go back at some point, they may not have a way to read that. In fact, odds are they don't because they have changed vendors or changed hardware platforms, and we can say they have an obligation to keep that capability, but the reality of it is they are not going to.

DR. CLUNIE: That is exactly Stefan's dilemma. What he is trying to highlight is if you define the source data as to be the data off the scanner before the site does a job deidentifying it, then the CRO can't retain that because they only receive post-source data material, if you like. So the CRO can retain what is sent, but the site, if it doesn't fulfill its obligations, then essentially the source data is lost forever.

DR. CHARLES: I think the only place you will ever be able to recoup that data in a reliable way is either at the CRO or if the sponsor uploaded a copy from the CRO because if you go back to the site some years later, it is just not going to be there. That is just the bottom line.

DR. CLUNIE: I think the two requirements are fundamentally in conflict. The requirement to preserve the source data with the patient's true identity is in conflict with the requirement not to have the patient's identify, and you can't have both at the CRO end.

DR. CHARLES: Right. But with respect to the trial data, that may not matter.
The other issue I think that gets back to some of the earlier comments is what is transferred at least in the United States really has to be addressed by what is in the consent form. You can transfer information in the HIPAA regulations, as you will know, but the bottom line, if the patient didn't consent to a particular identification information that is constrained by, say, HIPAA regulations, then your next level, you are constrained by HIPAA, but the consent form will trump the HIPAA regulations if it is properly worded and addressed.

DR. CLUNIE: I have always preached the same thing, but unfortunately, that doesn't work in Europe because patients aren't allowed to consent around what the privacy directive says.

DR. CHARLES: Yes, I know. That is why I said in the U.S.

DR. CLUNIE: Right. The second thing is we have HIPAA II to contend with. I don't know if many people in the room are familiar with the fact that Kennedy is back on the privacy horse, and there is work to extend the scope of health record-related privacy concerns, such that it may not just be what we currently construe to be covered entities who are covered under the new bill.

The third question is a matter of practicality. Many sponsors see the presence of personally identifiable information in the CRO or in-house as a risk, and to mitigate this risk, they are forbidding its reception, regardless of what HIPAA says and regardless of what the patient might or might not consent around.

The other thing is regardless of what HIPAA authorizes the site to consent, the sites IRB may elect not to export identifiable information, regardless of what the patient might consent to or what the sponsor might desire.

On every count, we are essentially heading down the track of not being able to receive personally identifiable information.

DR. DORFMAN: It is unclear to me from this or from the document whether we are heading down a route of saying that for some of these things, there is just not right answer, it is whatever you document. But is there at least a sense that you will do the same thing throughout any particular trial as opposed to allowing, for example, the diagram you showed where one site might do it this way, another site might do it this way, another site might do it this other way, but at the end, here is where we have got at the CRO? Is there at least the sense that that degree of variability within a given trial can be taken out by demanding that part of the participation agreement is that this is how it will be done in this trial? So at least for that particular set of data, you have constrained the variables to the extent possible, or is the sense that there is no right answer, and any of these things leads us to this end state which, while there is some variability, we are going to agree that that is acceptable since we can't really quantify it and that is good enough? That is the part of the discussion that is unclear to me at this point.

DR. CLUNIE: As you say, there are three actors in this equation. There is the site, multitude of sites. There is the sponsor, and there is the CRO.

Now, all imaging CROs essentially behave in the same way in that whatever they receive from the sites, they archive, and they further deidentify whatever they receive from the sites because even if the site has deidentified it, they have probably done it wrong and completely used the wrong form of the identifier, wrong number, padding of zeroes and so on.

To my knowledge, all CROs perform the same way. They get it, they archive it, and they further deidentify it. So, regardless of what the site has done or what the sponsor wants, they have that process, the one exception being when the sponsor insists that they not retain the information that was sent from the site, and we have a proportion of sponsors who say that.

They expect us to crunch of the CD and throw away the original digital files we read off it and only keep what we have cleaned further, shall we say, to make sure that there is no PHI leaking into our systems, which is extreme, but okay, fine. We could agree that that was the appropriate process across all the core labs. That would be a possibility.

At the site end, I don't think there is any possibility of standardization because you run the gamut from the single stand-alone modality in a clinic to a PACS system, to a regional health care enterprise where everything is integrated, to somebody in Eastern Europe who is printing out their bone scans and scanning them in as .jpeg files and mailing them to you.

I don't think we can ever expect consistency on the site. What we as CROs would hope from the sponsors would be to get consistency on the sponsor side with respect to expectations. That may be an unrealistic expectation, but if sponsors could agree that they either did or did not want a PHI to come from the sites to us, that would be a good start.

DR. DORFMAN: Right. So I guess the point I am making, because this really has to do with the talk on pseudo anonymization, deidentification, which is what I think we were talking about before, and archiving, that maybe the document could recognize that there are a variety of schemes, but that the standard to which we aspire through this charter would be to specify that in the charter, you state how it will be done for that particular trial.

I don't think it is reasonable to expect that all sponsors across all trials will suddenly come to an agreement that this is how we will do it for time in memoriam, but I think it might be reasonable to at least expect for any given aggregation of data, it will be done the same way.

The problem is quite similar to accruing sites, for example, who are going to administer an experimental therapeutic. You could make the same argument that at some places, it might be given by a nurse practitioner, and in other places, it might be given by an RN, and other places, it has to be given by the PI, another place, itself administered by the doctor, but most sponsors would say for the purpose of this trial, we are all going to do it this way.

If the site doesn't want to do it that way, then they are generally not invited to be a site. Why aren't those same standards applied to the aggregation of image data, I guess is the point I am making, to decrease that variability, and shouldn't the charter state how it is going to be done for this trial?

DR. DORFMAN: Your last point, no disagreement. The charter should state what is going to be done, if we think the charter needs to reiterate that.
The purpose of the charter is not necessarily to repeat the SOP. So whether it needs to be in the charter, whether it needs to be reviewed by the agency, because I don't think the agency really cares how we deidentify data, so long as what they get is okay. So the presence or absence in the charter is one discussion point.
The agreement among sponsors as to what their expectations should be, you are on a continuum where at one end, you have better quality control, and at the other end, you have less risk of privacy leaks. Where you choose to be on that continuum is essentially a business decision, not a scientific decision.

DR. DORFMAN: Understood. I am using the word "charter" loosely to mean the entire construct of the protocol, the charter. Somewhere it is either states and referred to in the other place or it is stated wherever it is, but my sense is that there ought to be a clear statement for a particular trial and at least our document on the charter ought to suggest that that is a reasonable practice.

DR. CLUNIE: Do any of the sponsors have anything to say about this?

DR. MOZLEY: David Mozley for Merck.
I think we covered this ground yesterday, but as you know, it is Merck's position that this kind of information goes in other documents, not the charter, and these documents could be variously named, like the imaging operations manual for investigators or what have you, but not the charter. Thank you. That is point one.
Point two, we recognize that these are enormous challenges facing the field. We are not sure we are ready to make recommendations to the FDA about how they should construct guidance on some of these policies because, as all of you have alluded to, when we are running trials at literally 400 sites at a time, which we most certainly are, all around the globe, we face challenges that we ourselves have not really learned how the Federal regulatory agencies can help us yet.
So I would just put that aside for a next follow-up meting.

DR. CLUNIE: To play devil's advocate, which I am known to do from time to time, we shouldn't expect the agency to hold our hands the whole way through. The agency has a certain purview based on the act, and patient privacy may not be their primary concern, as opposed to efficacy and safety. Human experimentation is covered by other regulations promulgated by other people, and again, only predicated on Federal funding. So there must be a place where as an industry we can have a good practice that we agree to that is not necessarily something that translates into a guidance from the FDA.
So I hope that we don't focus our whole effort on this document on pre-FDA guidance material, that we should be able to make statements that stand on their own.
Be that as it may, I am not sure that we can get consensus on this particular issue in terms of the sponsor expectation of how much privacy information to transfer from the site to the CRO.

MR. McCORMICK: Terry McCormick from IBM.
One of the questions that occurred to me is was there any discussion about the location of the data that is stored either at the site or the core lab, because I believe some people view that strict interpretation of the guidelines or rules is that it should be kept at the physical site of acquisition, and we know with the variabilities, building on Gary's analogy, there's a lot of things that happen at the sites.

A lot of large hospital systems migrate their information for archiving to another site that might be nearby, but it is definitely another address, so to speak, and do we want to think about putting into the document any kind of guidelines about principles about how to ensure either the physical location or the accurate reproduction or retention of information, regardless of location?

DR. CLUNIE: That is a really interesting point. The GCP statements on this and the CFR statements on this are old, and they sort of predate the notion that you might want to have a reliable off-site archive for backup purposes and the HIPAA security rule kind of constraints.

But in terms of return on investment, preparing this document, we picked the areas that were a source of pain to sponsors and to CROs, and what happens at the site and whether they will survive a site audit for cause or whatever, I don't think that is really high on our agenda as concerns. I think we have been content to gloss over such issues, not to say that we couldn't put it in the document if someone wanted to make some statements about it, but I don't know that we want to run around burdening the sites with extra stuff when they already have quite a high burden to bear.

MR. McCORMICK: Thank you.
I think, though, the existing rules, if they think that they have to retain it for up to 15 years or more, it is rather burdensome, and maybe there's some realities that could lesson that burden to them and the sites and the core labs.
DR. CLUNIE: I think in jurisdictions where there are such retention requirements for clinical reasons, the clinical retention requirements vastly outweigh any of our GCP-related requirements. So, essentially, we are flogging a dead horse there in the sense that they have other, much more pressing reasons to retain the information reliably, but I don't think you can generalize that internationally.
DR. CHARLES: The research data is not necessarily put into the clinical record.
DR. CLUNIE: That is a good point, but for registrational trials, which is our scope, the vast majority of images acquired -- and I am making a broad generalization here, of course -- are clinically examined, if only for reimbursement reasons if nothing else.
DR. CHARLES: Absolutely not. I sit on an IRB, and I can tell you if it is a research study, we explicitly state to the patient that the research component of the study will not be included in their medical record.
DR. CLUNIE: How does Medicare pay for it then?
DR. CHARLES: We don't do those.
DR. CLUNIE: Maybe you don't, but let's have a show of hands from the room. How many people have most of their trials with their clinical images as opposed to specialist research, paid for by the sponsor, images? I would

say 99 out of 100, if not 999 out of 1,000 in oncology, and I hate to pick on oncology.
DR. CHARLES: Ah, yes. You speak to hijacking the document. This gets to the fact whether it is a registrational trial by a sponsor or the most egregious violator of what I consider an ethical rule, and that would be the NCI that does not want to pay for the scan. So, yes, they expect third-party payers to cover these issues. Then you get into a different problem. That data flows over into the PACS system and is archived according to whatever the clinical rules are, and you have no control over that, and to your point, all of this is irrelevant.
DR. CLUNIE: For the sake of argument, there are two categories, and we need to address them separately. I think it would probably be a valuable addition to the document to talk about expectations of sites for storage of research-only, non-clinical data. That is gap that is not filled in the document. So, if you want to write that, that would be great.
DR. CHARLES: Let's make two IRCs, one for oncology trials and one for the rest.
DR. CLUNIE: That would be fine, too.
DR. CLUNIE: But if you want to write it, Cecil, we will be happy to incorporate it.
DR. CHARLES: You already have it. You just have to split that one piece out.
DR. CLUNIE: Any other comments on this subject?
DR. YOUNG: I just wanted to ask a question because a lot of the issues that I see that hit my desk are around site's ability to deidentify data and nothing quite rightly recognize that as a priority, a bit of my territory, but you alluded to a rework of the DICOM standard to facilitate anonymization. Was that right, David?
MR. BAUMANN: Yes. That is Part 15 of the DICOM standard, which describes how to protect images.
DR. YOUNG: Have you got any plans as a working group to facilitate that?
DR. CLUNIE: Yes, we were thinking of having a meeting at RSNA about the subject.
Working Group 18, Clinical Trials Working Group, met earlier this year to discuss what our agenda should be, and there are a couple of pressing issues, and this particular group had to come to its conclusion first to provide the input for Working Group 18 to do the work.
Now, having said that, it is easy to make a list of attributes that are risky, but then there are many good reasons for preserving some subset of those attributes or some information that is in them.
Correct me if I am wrong, Stefan, but I think the objective here is to say if you as a device conform to this profile, then that means that you have removed or made removable this set.
The purpose of the DICOM standard is to allow interoperability between different devices and to make sure things comply at a certain level. So writing it in the DICOM doesn't mean PACS or modalities or work stations are going to either adopt the profile or claim compliance to the profile. It serves more as a central place for the technical people to discuss the matter and then implement the proprietary end or free software to do the job.

DR. YOUNG: That was the other thing that I was going to come to was that sites often, apart from finding it tricky around deidentification, don't have the right sort of software or have issues with their software. Is there a plan within the community that collects images for pharma to supply versions of appropriate software to do this, or do you do it already?

DR. CLUNIE: A couple slides later on will be talking to the general nature of that issue.
Sometimes we supply software, but sites don't often want to run software that other people supply. The clinical trials' guys may have one agenda, but the local IT guys have another, and so it is not necessarily easy to supply the sites with software. There are many web-distributed solutions that address this particular class of problem, and it may just kind of go away as we move to Internet-based transference instead of CD-based transference. That is the hope, but certainly, it is not uncommon practice to supply them with a decent tool that you have validated to do the job. Getting them to use it is another matter.

DR. FROST: As an investigator and an IRB member, I would concur with the comments of the gentleman at the rear microphone about the number of research scans only that often come into play, and I think we are going to see an increase in that as we look more and more at mechanism biomarkers where maybe you do a baseline FDG PET in the treatment, but then you have other research scans interspersed within those. Oftentimes, those don't really go into the clinical flow either. So they are not interpreted by the radiologist because they are not reimbursed. There is not a report generated. So this traces back I think kind of to the duty around the viewing and screening of those scans as well that we talked about later on.
So I think this is going to be an increasing issue as we look more and more at mechanisms around drug actions, and it will need to be addressed.

DR. CLUNIE: I think one of the things that hampers what you are describing is the fact that clinical PACS don't provide provisions to help researchers. They don't provide ways to make it easy for researchers to store their data and segregate that data from clinical view. There are some PACS systems that basically, if it is in there and there is no report on it, it keeps nagging you to provide a report, and then sends out a bill for it, because they are focused on commercial reimbursement, not on storing information, but that is gradually changing, and I think it is our burden to try to convince the device vendors.
Is Luna around here somewhere?
We will make GE, for example, implement decent research facilities in the GE PACS to make your life easier, because it is crazy to have a researcher come onto the same magnet that is used for clinical care, run a research pulse sequences or one extra research series and somehow have to store that somewhere separately from where all the rest of the information goes. That just doesn't make sense.

MR. RAUH: David, could you expand on the archive idea of being able to get rid of that original copy that you receive at the core labs?

Lilly has taken a real strong stance on trying to minimize what personal health information get sent to the core labs, but it seems like we have been constrained around that CFR Part 11 and the whole audit trail thing where you have to keep a copy of what you receive, you further anonymize it, and so you still have that original copy. You made mention that some of the sponsors basically destroy that. I don't want to go against that because I think that is where we would eventually like to get to, but I thought there was a Part 11 type of compliance with an audit trail issue of data, or is that not the case anymore?

DR. CLUNIE: Certainly, it is always a dilemma in clinical trials to figure out when Part 11 is triggered, both from a validation perspective and an audit trail perspective and an electronic signature perspective.

I think the CROs perceive that it is their responsibility -- and correct me, CROs, if you disagree with this -- that we are only responsible from the point of reception, that we can't control the sites.

If we instrument the sites with, say, a web-based tool to submit the data to us, then we are responsible to some extent for that, and I think that is the dilemma that Stefan was referring to where the deidentification is performed in that tool, maybe by the applet running at the client's site, for example, and then if that produces an audit trail and you send it to the CRO, then you have defeated the purpose because the audit trail says Mary Smith was turned into 1234. So you can't have that if you want to attain your goal of no PHI at the CRO.

Then we addressed the issue of, well, should the site retain an audit trail, and we know that it is difficult to instrument them from that perspective. So the expedient solution, which you might have caught on one of Stefan's slides, is paper, printed out and tell the site coordinator they have to keep that forever with the rest of their trial material -- I say forever -- two years with the rest of their material.

That is arguably an expedient solution to dealing with the issue. The current standard of practice, as far as I can tell, is to do nothing, that the site does whatever it does, and when we get it, we start out audit trail from that point.

Now, in the case of, for example, one of your trials and one of your patients, if the site has failed to deidentify it, as you have stipulated that they should, and it really has the patient's name in it and we deidentify it, then should we keep an audit trail of that? Because again, that kind of defeats the purpose.

I think it is unrealistic to expect perfection, that there will always be PHI at the CRO, and there will always be PHI at the sponsor because someone is going to send you an e-mail with the subject's initials in it, for example. It is just going to happen, and it is not like you can throw it away, but we can strive to avoid it, I guess, and I think that is the best we can do.

But I believe the CRO's job is to start an audit trail from the moment the material arrives, and I can't see any way around that.

You are shaking your head there.

MR. RAUH: Ideally, we don't want you to have that information, but we understand the kind of Catch 22 of in cases, in many cases, in some cases, you are going to receive some of that information, and as far as I know, all the core labs have a process to anonymize that. So you are kind of caught in that Catch 22 of the audit trail. You are required to maintain an audit trail, but by maintaining an audit trail, you are also maintaining some personal health information that we may not want you to have.
I didn't know how much of a dilemma that was and collectively what's kind of the thought around that. Is the audit trail more important, or is getting rid of the personal health information?

DR. MOZLEY: I think the key question is what portion of the audit trail is important, and here is an area where the FDA could help us.
If you imagine a worst-case scenario where the FDA finds it necessary to conduct what we could loosely call a hostile audit, what would you need to view in order to reconstruct a trial and verify the veracity of the key outcome measures?
I would suggest that you could start with a deidentified image, and that would be source document one. I think that the marked-up images also have to be classified as source documents, and there has to be an audit trail between the first source document that is not marked up and the final marked-up document.
Then the outcome measures can be described as scalers or verbal descriptors or the like. They are the outcome measure.
So the key point is what is it that needs to be audited in this scenario, and Merck's position I think would be -- I think I can speak for the company -- would be that we need to start at the point where the FDA inspector would start, and that would be at the deidentified image.

DR. CLUNIE: I think in summary, we should have such a discussion and try to achieve a definition of what an audit trail is, what source data is. So it is open for discussion if we can talk with the agency about what that means.
However, remember the reviewers are a different breed to the compliance auditors, and if you watch your GCP auditor at work, then one of the things they pick on is inconsistencies, and inconsistency of identifiers is a key because they know from experience that maintaining identification is difficult.
If you are auditing a site, same issue, and there is also a matter of quality, independent of satisfying an audit, in that if you want to improve your own processes or monitor your own processes, then you have to measure these things. You have to track these things. So, if you want to track the effectiveness of your deidentification process as a quality control matter, then you need to record what was before and what was after and look at the differences and see where there are discrepancies that need to be corrected.
So, if we are going to improve the industry, quite apart from maintaining regulatory compliance, I think we have to keep track of what we do.

DR. FORD: Thanks.
I just wanted to point out that this entire two-day effort is really on the basis of harmonizing imaging core labs and the processes by which they work, but I think a good partnership is usually a two-way street, and I think this is a good opportunity to maybe ask our pharma colleagues, that, in fact, there are opportunities for reverse harmonization or allowing pharma to take back to their organizations topics that we bring up as issues, and perhaps we can encourage them to have some harmonization on some of those topics that would then benefit the partnership in both ways.
So, David, I ask you to perhaps bring back to your organization, some of the topics that we identified that need to be harmonized on the sponsor side.
DR. DORFMAN: That was the point I was trying to make, whether it is in the charter or in the SOPs, whatever it is going to be. There ought to be some uniformity on this.
If as a community, one of the things we want to do is validate imaging biomarkers, having this amount of noise in there, whether it is through the anonymization process, how it is transfered, how it is retained, how you build these archives, it is not going to be to any of our benefit.
The other point I was going to make is for that RSNA meeting. I applaud that, but, David, as you know, the fields where folks admit to having PHI is not so much the problem. It is those fields that are kind of vaguely defined where we are not sure whether there is or isn't PHI, but there also may be information that is useful or potentially useful in the downstream analysis. We went through all this with the CA big stuff, you know. How about those fields where we are just not sure if they are proprietary, corporately defined, use fields, and we don't know what is there and what isn't? So we strip them, but could it be information that might be useful in the later analysis of those data?
So, hopefully, the flip side of that is not only to say where PHI is and, therefore, it can be safely dealt with, but to have some places that are clearly off limit, even in those user-definable fields.
DR. CLUNIE: Absolutely. I think that is actually what the basic application confidentiality profile is called, PS3.15. It tries to address by saying here is everywhere that is at risk, but do you remove study description, because that might say something very important to you, and it may not be said anywhere else, given the vagaries of the vendor's choice of encoding information.
But it is hard to mandate a strategy that is universally successful. So, for example, if you just extract body parts from study description and just keep that, then what happens if it is Mr. Hand or Mr. Foot?
DR. CLUNIE: So, no matter what strategy you apply --
Seriously, I have done a lot of work on this.
This is worse on an international scale where you have the French word for hand or the German word for hand or the Portuguese word for hand or the Estonian word for hand.

So the tooling is important. It is not a deterministic process, and there will always be some fields where there might be something useful, but there is just too much risk, and there are other fields where there is likely to be something useful, and the risk is relatively low, but you will never achieve 100-percent deidentification, unfortunately, no matter what strategy you take.

I think this is one of the most important topics for Working Group 18 is to subcategorize the fields into the sort of categories you described, what is really risky, what must be removed, and what is sort of slightly risk that bears attention, but you may have some small amount of leakage through.

MR. BAUMANN: I think to expand on that, the approach that we are taking is that if this should ever propagate as a standard, we have to say by default, "It is bad and don't collect it, but if you see specific reasons, do pattern checking or collect it, if you are sure enough that it doesn't contain anything."

DR. DORFMAN: I was intrigued by your comment about it is crazy that you can't put research images into the clinical PACS system because we think it is actually crazy to do that, and we have created at Duke a research image management system. So the data that goes in the clinical PACS goes through a work process. There is a read, so on and so forth.

Research data goes into a protected account for that project, so that only key personnel associated with that project have access to that data per the common rule, per the IRB process.

So I would argue, in fact, that you shouldn't be putting research data into your clinical PACS system because you have really sort of violated the intent of isolation of research data from all but the key personnel associated with the project.

DR. CLUNIE: Oh, there are definitely many different ways to skin that cat, but have you ever worked in a hospital where there were two cardiology groups who won't let you allow the other to look at their patients' images? Then you know that even within the clinical PACS, you have to segregate information.

Most large enterprises are going towards having central IT control of archive for everything, be it lab data, be it radiology images, cardiology images, EKG data, EEG data, whatever. So while what you say is true, it presupposes the existence of a separate infrastructure that must be maintained and supported in a reliable data center, off site replicated, to protect it from fire and all that kind of stuff. Most enterprises just find it easier to manage it as a whole.

You are lucky enough to be in a university environment where you have that kind of luxury, but in traditional clinical trials for registrational drug environment, much of your imaging is performed in community hospitals, clinics, all of whom don't have such ability to build a special infrastructure just for research.

I don't disagree that is not a valid paradigm, but I think we need to account for both.

DR. TOTTERMAN: I would like to echo the concerns of a couple of previous speakers. I am afraid I have a concern that if we go deep into it, to try to eliminate all the information, patient-specific information, at the same

time we are eliminating information, what is very critical to run the clinical trial.
I was looking into the fields, what you suggested that contains patient-information, kind of risky to have there and some others, and I feel it is very crucial, have very crucial information. So I just warn about that, a concern about that, that we don't go too much into detail for the sake of HIPAA.
DR. CLUNIE: I think Stefan highlighted your concern in the sense that for some applications, it is appropriate to remove everything, but there will be special profiles for a particular kind of PET acquisition, for a particular kind of acquisition that is of concern to you. So, for any particular purpose, one can generate a profile that is appropriate, but the kind of trials that you do, when other people do them, you should have similar needs, and if there is enough value in defining a common profile for that, then we should do so and relax the kind of constraints that you have reaction to.
DR. TOTTERMAN: Right, because it is crucial for any MRI studies to really know what system was used for the study, up to the software upgrade level to understand if there is something wrong where the reason might be and also just particular information against what we do have.
DR. CLUNIE: Oh, that is a very important point. I think in generalizing the detail in the document, mention of removal of equipment information didn't mean take out the fact it is a GE machine or that it is a GE Signa or that it is a software version 1.2.3.4, but rather that the device serial number be removed.
Now, you can argue that in a particular trial, it is important to know the individual device, and when that argument is made, you should retain that attribute, if you can rely on the vendor to actually populate that attribute, which is another question.
So there is a continuum there, but the intent is to remove the individual identification of that particular machine, so that a technologist from GE can't look up their database and say, "Oh, this was done at Memorial Sloan-Kettering in Suite 2 on such and such machine," for example, not to say it is not a GE machine or a software revision. Does that address your concern?
DR. MOZLEY: Can you clarify that, please?
DR. TOTTERMAN: Yes. That actually is a concern.
DR. CLUNIE: Yes. There are some situations when you want to know the device, but in most situations you don't because it can be traced back.
DR. TOTTERMAN: One of the reasons I am concerned about that, in some of the studies, the studies have to be done exactly in the same system, and to be able to follow that the system actually was used for the information, the information needs to be somewhere there.
DR. CLUNIE: But the problem with that is that you are assuming the value doesn't change over time in the same scanner. If you get a software upgrade, they change the device serial number, or they don't populate it at all because these are optional fields.
So you can't rely on that information, but I agree that there is value in keeping them. Again, this is one of those tradeoff issues. The highly unlikely scenario that someone is going to track that device serial number down to a

particular machine, as opposed to its value in a clinical trial -- and anything in this document here as input to Working Group 18 is subject to discussion. So, if we don't have consensus on the equipment identifiers, then we should make sure that you are involved in the subsequent discussion on that.

DR. TOTTERMAN: Okay. The next one actually is related to the identification at the sites. Somehow I feel that I am responsible for the sites and responsible that they do not only the studies and imaging in the way, how they are supposed to do, but also they do the deidentification in the way I am expecting them to do.
So, if we can train the sites in the way that they actually follow the rules and regulations, that will make our life much, much easier because images and studies will be deidentified before they come to us. Just a comment.

DR. FUERST: I have some comments on the same topic. I think, first, we have to make sure we are thinking differently about HIPAA and PHI versus information that we collect for a clinical trial. There is PHI and restrictions about information that identify the patient with more degree of geographic precision than the zip code or something like that, and a scanner serial number could do that, but in the sense that a patient has consented to be in a trial, they have consented to provide certain information, and that might allow identification of their location more closely than a State or a county or a zip code, and I don't think that is necessarily a problem.
It might become a problem if that same scan is then going to be provided in a public release form, and that serial number might now be able to be used by somebody to link back. That is something different, but I think we have to keep those things separate.

The other point is that I get a little bit concerned when we deidentify too much because we know that sites will make mistakes. If the only thing we collect is an image and an ID number, there's going to be mistakes there, and we are going to get the wrong patients associated with baseline and follow-up images, and it is going to be kind of a mess, and I think we have to have certain links that allow us to go back and to have that kind of connection with the source data, so we can verify that. I am just a little bit concerned about that.
I am uncertain about to what degree we have to worry about reconnecting images with information that should be kept private at a hospital. If you have a UID and that is all, is that really sufficient information to go back and to find out who that patient was? Theoretically, you could if you could go to that scanner or that hospital and use that UID to find the source record at that facility. You could find that patient's name, but is that reasonable to expect that that is a risk? Because you have to have access to other private and protected information that we don't have control over.

Then to that extent, you could remove the UID, but then the image itself is an ID that makes it a little bit harder, but you could go back and you could match an image with an image on an archive and say, "Aha, I found this patient, and now I know their name and their phone number and their address," and whatever.

So it is degrees of what kind of separation, what kind of difficulty, and I am not sure where that line is that we have to protect ourselves against. I think that is a difficult point for us in clinical trials, but again, the PHI issue I think is a little bit different from what patients consent to for a trial, and PHI issues may become more important in a reuse or data release public repository situation versus keeping things controlled and restricted access within a CRO or within the context of a trial itself.

DR. CLUNIE: I think you have made a number of points, and just to make sure that we cover them in the document, first of all, the question of consent, if we are going to do international trials, we have to accept the fact that patients can't consent around certain regulatory concerns.

As we said before, consent works in the U.S. Consent doesn't work in Europe, sort of in a very general sense.

Second, with respect to UIDs, I agree that UIDs are great for sort of retrospective analysis back to the site, and I personally prefer to keep UIDs because I see the risk as so close to negligible that someone will take that UID, go into the PACS at one of the sites, and they don't know which site to go to, match them up and try to establish concordance, theoretically that is a risk, depending on where the threats come from.

If, for example, you had someone who was malicious and wanted to jeopardize a trial, was at a site, they could troll through all the identifiers in the dataset for the trial and match them against their own UIDs, identify this patient and say, "Hey, naughty person, you have not deidentified this properly. So give me lots of money." If one is going to be truly paranoid, it depends on the kind of threat we are trying to protect against.

Doing a serious hazard analysis is just too complicated in most cases. So you have to come up with simple generalizations that will satisfy the IRB, and the safest thing to do is simply to remove the UIDs, without any question, even though they have value in a retrospective order, but it is certainly safer to remove the UIDs.

Further, most sites will run deidentification software on the images that will remove the UIDs anyway and not keep an audit trail of it. We found that the UIDs from the sites are generally unreliable from an auditing perspective and also generally malformed or not unique and, hence, have to be coerced into new proper values anyway. So UIDs unfortunately lost their uniqueness and their value.

The other thing you mentioned was with respect to the reused dataset as opposed to the dataset used within the context of a trial. I think that is a very good point, and we should specifically develop a profile of addresses to reuse use cases, which is much more aggressively deidentified perhaps. It might be necessary from data we are sharing with readers who are subjected, you know, professional constraints, ethical constraints, and contractual obligations.

DR. FROST: Just to continue the paranoia, you could also have cases that are completely deidentified, but you have a rare tumor. Say a 52-year-old man in an institution with a paracardial sarcoma, something very rare, there is only one of those. So, if you have that information, you can theoretically trace it back to that individual. When I was on the Hopkins IRB, that came up a number of times. When the data gets distributed across multiple sites, clearly it is more difficult, but another level of paranoia to keep you up at night.

DR. CLUNIE: Oh, it gets worse than that. That is true from a clinical perspective, but if you actually do a hash of

the pixel data, then you can very quickly match an image in a clinical trial set to all the images in your PACS and find out who the patient is because if you haven't lossy-compressed the image or added noise to the image, then the hash value of the pixel data itself is even better than a unique identifier. So there are many strategies where you can aggressively try to reidentify patients if you have access to a set of known data about that patient.

This is the essential dilemma in deidentification, do you protect against people who are in the real world who have no access to patient information, or do you protect against the reidentification by someone who already has sufficient information about the patient. Those are really two different scenarios.

I would argue it is impossible to protect completely against the latter unless you go so far as to add noise to the image to prevent this use case. Then it becomes a statistical problem. The more noise you add, the lower the probability of a match, but you don't eliminate a match, and hence, you have to add so much noise to the image to degrade it, such that you won't get a match, that it becomes unusable, ergo it is not a sensible strategy, ergo you can't protect against this concern, ergo you shouldn't worry about it.

DR. ASHTON: I think I would like to echo that. I think it is very important to protect against inadvertent release of patient identification, but if we go to the evil genius scenario, as David aptly pointed out, there is nothing we can do about that. Descarte demonstrated 400 years ago that if you assume that there is an evil demon who can control everything in the world, you can't do anything about that.

So, if we assume somebody is going to have complete access to the PACS system, access to the data, really wants to find out who these patients are, I don't think there is a strategy we are going to put in place that is going to defend against that.

DR. CHARLES: That is exactly right, and you should say one more thing. Those people are violating criminal statutes in doing that, and if they get caught, not only will they be fired from their institution, but they will likely go to jail and meet a guy named Bubba.

DR. CHARLES: I think it is really critical that people need to understand, to the point both of you made. We can't protect against criminal behavior. It is a ludicrous concept to think that we are going to come up with some structure that is going to prevent us from a criminal going in and doing this in a malicious manner. It can't be done.

Again, we could completely destroy the data, to David's point, but then what would have been the point of the study in the first place?

DR. CLUNIE: Anybody have any final comments?

DR. CLUNIE: We will let Cecil have the last word.

So, on that note, since it is only five minutes until the break time, I think it is better if we stop, have the break, and we will resume the other topics for discussion in the next session because I don't think we have time to do justice to the oral presentation.

Session 4

Data Integrity and Statistics Analysis Plan Requirements

Panel Discussion Leaders

Dr. Edward Gastineau, Chief Executive Officer, ICON Medical Imaging

Ms. Wen-Lin Luo, Statisician, Merck Research Laboratories

Dr. Jyoti Zalkikar, Mathematical Statistician, FDA

Dr. Rajeshwari Sridhara, Deputy Division Director, Office of Biostatistics, Statistical Team Leader, Oncology Drugs Division, CDER, FDA

MR. GASTINEAU: Before I get started, I would like to acknowledge over the last couple of days, I think it has been a tremendous meeting with a lot of great content, a lot of great discussion, and one of the people I think that needs serious acknowledgement is Constance. She has done a tremendous job over the last few months kind of shepherding the cats, as it were, and getting us all onto the same pages and arranging teleconferences and meetings.

So, if we could, give Constance a big round of applause.

MR. GASTINEAU: Well done.

Over the next four, maybe five hours, we are going to go through stats and a serious number of equations and things and some of the consensus that we reached.

So we will skip that part, as I am not a statistician. I don't play one on TV, but we are surrounded by folks here who I think are going to be able to help us with a lot of the discussions and points.

Some of the basic background and considerations that we tried to take into our working group was to look at the stat plan from a generic approach. The stat plan, as I think everyone acknowledges, is a separate document. We acknowledge that each of the therapeutic indications have content specific to each of them, and we wanted to keep this more on a generic level and sort of a methodology that would kind of fit into the development pieces.

Our focus was on Phase 3 and late Phase 2, so, again, consistent with the IRC approach, and we were assuming that these are based on accepted imaging practices, such as CTs and bone scans for the use in response and progression monitoring in cancer trials or x-ray and MR and osteo or rheumatoid arthritis. So those are the type of situations we were working under.

In terms of a basic overview or premise that we worked from as well, the development of the content and the consideration for the SAP is referenced into the IRC.

I guess the first contentious point perhaps starts here, and it is really about the continuation of the conversation

we had about how much information from a separate document, i.e., the stat plan, goes into the IRC.

In Working Group 1, if we looked at the table of contents that was proposed there, what we see is that the stat plan is referenced as an appendix in the table of contents there, and the questions, I guess, that sort of are on the table here, we acknowledge that the stat plan is a separate document. It is associated with the protocol as the primary documents that guide the study conduct. Those are the study documents that typically get on to the agency for review.

The charter is one that fits sort of at the end of that process as the implementation plan, if you will, of the derivative of the protocol and the stat plan, and the IRC at the appendix level would contain a summary and a reference back to the actual working SAP.

So that is the content by which we are working under, and some of the description may be around how much content goes into that summary. There is some differences between agencies, divisions. There's differences between participants in this level and how much should be referenced there, how much redundancy should be in place, and I am sure that will be a hot topic of discussion.

Some of the topics that we hope to include over the next couple of slides are the process and timing of the development of the IRC and the stat plan, on site versus off-site reviews, and we covered those briefly yesterday, design metrics and consideration, what goes into the thought processes around this, how do we handle missing data, adjudication, blinding, reader performance. So those are the key topics that we have tried to touch on and hopefully will have some more discussion on today.

One of the things we talked about was the chronology, and I think over the last couple of days in terms of the IRC piece, certainly the intention has been that the IRC be a dominant document within the context of protocol, state plan, and IRC as a submission piece, and there has been discussion about where does this come into place.

I think from a practical perspective, what we have seen historically is that that IRC needs to be done a priori. It needs to be in place before any analysis or blinded reads or something of that matter start, so that those processes are defined.

I think perfect-world scenario, we are all sort of in agreement that the IRC should happen before the study begins and be part of the submission piece to get comment back, but I think the underlying rule, if you will, that we are working from, from a practical basis, is it needs to be in place before the analysis starts.

So we will get into the on site versus off site, and these are just bullet points for discussion. Hopefully, my statistical colleagues here will help us sort of go through some of these pieces.

In terms of just setting the background here, on site, off site, we had some of those discussions yesterday. The sort of standard today -- and many of the indications is that the -- I think I have made a typo there. Off site is the typical primary dataset by which we are looking at efficacy as a way to minimize bias and reduce the outcomes for that.

We typically think about the on site and the off site as two separate datasets. There certainly are comparisons or discordant analysis that are conducted between the two, but in general, they tend to be monitored as two separate pieces. So, with that said, I would like to open the floor to the table for comment and discussion around perhaps the FDA's view or Wen-Lin's view on, on site, off site.

DR. SRIDHARA: Good morning, everybody. I am Rajeshwari Sridhara I am deputy director of the Division of Biometrics in FDA. I am not in the imaging team, but on the drug oncology team. Of course, we do have some of the imaging endpoints that we use, such as tumor progression or tumor response.

With that, where do we see this on site and off-site reviews? I think we do understand that there will be some differences between the two of them. The question is how can we reduce this or why does this happen.

Probably, if we look at why does a difference happen between off site and on site, in our experience we see that it is not the same lesions that the two are seeing. So that in itself will bring in some differences.

So whether one part is prior to the start of the treatment at the time of randomization, whether you can identify the target lesions or both the off site and the on site to follow the same lesions, so that at least you have the same lesions that you are looking at, the other issue has always been that the on-site investigator is not only looking at the radiological data, but actually listening to the patient, and there could be some other things that are happening to call it a progression and move on and change the therapy at that point, particularly in oncology.

So, with that, the direction of the treatment has changed, and if this is the efficacy endpoint that you are looking at, then what does this independent reviewer assessment mean in the overall evaluation of the efficacy endpoint? That becomes an issue. Then the question is how much is the discordance.

If you ask me would 10 percent be acceptable, would 30 percent be acceptable, I cannot answer that question. It truly depends on the disease that you are looking at, and whether we believe that what you are measuring is actually true.

To give you an example, this is a statistician speaking. So bear with me. I am not a medical person, but what the clinicians tell me is in brain tumors, if they have had radiation as one of their treatments, then it is harder to read their scans. So whether you are actually reading the tumor progression or some other inflammation, it is difficult to understand.

So, in such situations, it may be that we may not even accept a radiological review of it, although the primary physician may be using this as a tool to see where the patient is going. So we are not saying that in practice this won't be done, but for the purpose of evaluation of a therapy, this becomes a problem.

Statistics can help only if you have actually the correct measurements that you have. So, if you don't have the correct measurements, then it becomes a problem.

So, in some diseases, it is difficult to measure, and that is what the clinicians are telling us. So it is not coming from statisticians that, "Oh, you have this huge discordance, so we throw it out," but it is coming more from the clinicians as well, where they feel that in this disease, you can certainly use a radiological measurement for

assessing the disease.
So there is no firm way to say this is the way we assess it or not. So it does very much depend on the disease and what stage of disease and so on and so forth, and also it will depend on what is already available therapy for that group of patients and so on.
So, with that, I will ask for Jyoti to add if she has anything.
DR. ZALKIKAR: No.
DR. LUO: I think from the statistical point of view, we probably need to specify early on in the protocol developments. From my experience, in the SAP, we have to specify early on in the protocol development stage what is the primary endpoint that we are going to analyze.
For example, either on site or off site, then we specify that in the protocol or in the SAP. At the end, we probably will have two set of data. One is on site. One is off site. Of course, the statistician can do kind of the discordance analysis and provide that to the agency, but we will kind of make the conclusion based on what we say is the primary in the SAP, but I don't know about other companies. The statisticians can also have some input here.
DR. SRIDHARA: I just wanted to add one more thing. We are talking about discrepancy between off site and on site, but we have also observed that even off site, independent reviewers themselves have discordance. That will really bother us very much. These are supposed to be experts reading the same things. So you would expect that there will be very minimal, if at all, any discordance.
If there is discordance, then the question is whether it has something to do with the disease itself that you really cannot measure properly. So it could be that in such a disease, if you do find discordances among independent readers, then it is more of the disease issue rather than the readers having a discordance.
Again, it is very much dependent on the disease that you are looking at and what lesions are being read.
MR. GASTINEAU: Gary?
DR. STEVENS: I would like to address the whole issue of discordance. My real question is why because we are actually measuring two completely different things here.
Concordance and discordance from a statistical perspective was set up to assess measurements made on the same experimental units. In this particular situation, we don't even have the same experimental units. The on-site reader is looking at a patient and images together at the same time. Whereas, the off-site reader is reviewing randomized images. So the experimental units aren't even the same. So discordance in this particular study doesn't even make sense.
So the real question is why are we even worried about it.
DR. SRIDHARA: If I may ask, what do you mean by randomized reading?
DR. STEVENS: We talked about it yesterday. The images for these patients are put into a big pool, and they are randomized out to the readers. The readers are blind at the time to the patient's gender, sex, and all these other things. They have no information on what is going on with these images.

So the images themselves are the units being randomized to the readers for the off-site assessment. Whereas, the on-site assessment, the patient is the only thing that is randomized. So you are randomizing two completely different units here. The experimental units for the on-site reader and the off-site reader are completely different. You are assessing two different things. So discordance doesn't make a lot of sense in that situation.

DR. SRIDHARA: Yes, true. You expect a certain amount of discordance simply because, as you said, that is sort of saying that you have different lesions that you have selected and whatnot for the same patient, and so you will expect some differences there.

DR. STEVENS: I am going more fundamental than that. I am going back to the experimental unit that you are looking at, not lesions or anything else.

You have patients being randomized versus images being randomized. So you are experimental units are completely different.

The whole concept of discordance needs to be completely redefined in that scenario because, typically, concordance and discordance are assessments made on the same experimental units, and we are not even doing that here.

DR. MARZELLA: I think I would like to address one portion of a question that was raised, which is why does it matter, and is it really an administrative issue?

I would like to throw in another example where it is very important, and that is if you have an imbalance across study arms in the level of concordance or discordance that there is.

The other concept is the performance is not similar across arms, or the performance is not consistent with what is expected from the literature. Then our concern is, is there some sort of bias involved, particularly in trials where blinding or other features, randomization, don't work quite as well. So that is the concern, how do you interpret data where there seems to be differences between what the expected and accepted performance is and what actually occurs in the clinical trial.

I think that this issue came out yesterday in terms of what level is reasonable to expect, and it is more how predictable this effect is, this level of discordance.

There's also ways by which this issue can be addressed. For instance, even in the blinded read, there can be progressive unblinding. So that if you are concerned about the effect that some clinical information might have on the actual read, you could also in some way look at this issue.

So I think I would like to sort of agree with the colleagues in statistics who are saying that it is an issue that we look at, and again, the concern that we have is to try to figure out whether it is reflective of potentially some bias or some problems with actually study conduct.

Another example that came up yesterday, for instance, it could be some sort of outlier, sort of performance, which is related to a specific site. So all of these issues I think are important in the review.

DR. SRIDHARA: I just want to go back to the idea of this experimental unit being different. Maybe the off-site readers are getting the radiological scans in a different way, and they are assessing it, they are not looking at the

patient, but just the scans.
However, when we are evaluating a therapy -- I will go back to the oncology example. It may not be the case in the other diseases, but if we are looking at time to progression, it is that data where a reader calls it a progression is what we are looking at.
In this case, we have made it clear that it is the radiological progression that will define progression for the patient.
So the on-site reviewer will also have read the radiological scans and come to conclusion that this is the date of progression, and if that differs from the off-site reviewer, then which one is the correct one to use? Because we are basing, we are evaluating the therapy based on the time to progression.
So, even though you may not be randomizing, randomly giving the patients to these independent readers, it is still the patient data, and it is still the whole thing that you are looking at.
That is where we are looking at why it is for this patient, what do you call as the progression date. Do you use the off site, or do you use the on-site data?
In theory, one would imagine that if a patient is progression, it should be the same, no matter who looks at it. So that is the problem we are having.
If you are looking at, for example, survival where death is the event, there is no question of discrepancy. You have a death date, and that is a solid date.
Here, you don't know which one to pick, and that is the problem.
DR. RAUNIG: It seems to me that the fact that you can't bring up a specific rate of discordance, whether acceptable or unacceptable, it means in general these analyses are very difficult to interpret.
I have been involved with trials where you see discordance rates in excess of 30 percent. Yet, the hazard ratio when the PFS endpoint is derived from the investigator assessments or from the assessments from the independent review facility are very similar and both statistically significant.
I appreciate the concern about potential unblinding, but the independent review facility would be blinded. It seems as if you come up with an analysis based upon investigator assessment, which is significant, and you have another analysis based on an independent review facility, which gives similar directional results and perhaps also significant. I wonder why there is a need to do these discordant analyses which really just lead to I think unreasonable concern about the interpretation of the data.
So I would appreciate the agency's comments on this.
DR. SRIDHARA: I think the problem is -- you are right. We could have either just on site or off-site reviewer. At least in oncology, most of the trials are open-label studies. So there is a feeling that the on-site reader may be biased and knows which treatment the patient is getting, and therefore, they may be calling progression sooner in the control arm versus the treatment arm.
So, to make sure that that is not the case, we say we need an independent reviewer. However, for the practice of medicine, the investigator on site has to look at it at the time the scan is taken or whatever the regular visit,

and they have to move on with the decision-making.
So no matter what, they are going to read it, and you have that data. What we want to see is was there really a bias anywhere in this, and if there is truly a huge difference between the independent and investigator reviewer, then we have to question the reliability of this progression data at all.
As I said, it is different when you are practicing medicine and how you use it, but as regulators, when we have to see if a drug is working or not, then it does become important to look at any imbalances or anything that is happening.
DR. RAUNIG: But if the independent review facility is blinded and coming up with a significant result, the goal of the trial is to determine whether or not the therapeutic agent is effective, and if both analyses are showing that it is effective, what is the purpose of looking at these discordant analyses? It is not going to change practice of medicine as far as how investigators will decide to make treatment decisions.
What we are trying to determine is whether or not the drug is effective, and if you have a blinded radiologist, based on their data, determining that the drug is effective, I don't understand the additional value of doing these discordant analyses because there will be discrepancies, as you mentioned, if the site is unblinded and the facility is not, but if both are showing that the result is significant with the same directional effect, I think that is conclusive evidence that the drug is effective.
DR. SRIDHARA: I don't think if you look at our reviews when we are evaluating therapies that we are going into detailed discordance analysis, but we want to make sure that we can explain the discordance that is there. If we cannot explain, then we don't know even if we want to believe the independent reviewer.
So it is more the question of can we explain the variation that we are seeing and what it is, knowing that, yes, there will be some amount of discordance.
DR. MARZELLA: I think that the issue is not that we want to focus on just one metric. The example that you brought up is really useful.
If data looked at by different ways are trending in the same direction, then that makes us comfortable.
So the issue is not that we want to focus on a specific metric and a specific threshold for saying this is a problem. It is just that we want to look at these data in different ways and determine whether the data are believable.
The issue is also how unexpected -- we expect that there is going to be more concordance, we hope, with the blinded independent read, but the issue is we have sources, as we have heard yesterday, sources of variability all along, and what we want is we want the scientific quantitative way of looking at all of these measures of variability and trying to understand how predictable are they.
Again, the focus is not on a specific metric. We don't have a specific threshold, but we would like to take a look at all the data looking at different ways and seeing how consistent it is.
As I said, the other issues are not just globally, but we are also concerned about performance by site. It would concern us if there is a site that is enrolling 90 percent of the patients, where the performance is very different than at other sites. We are concerned if there is imbalance across study arms because of potentially problems of

unblinding. So it is very complex, and we are not just focusing on a specific metric and saying you got to meet this threshold. We are saying that this is an important aspect that needs to be looked at.
MR. GASTINEAU: David?
DR. RAUNIG: While the issue of discordance and adjudication is complex, you can initially set a lot of these issues down in the initial study design with some knowledge of discordance. We should have that knowledge somewhere in the literature.

We should have some a priori knowledge of what the discordance is, and that is going to be subsumed in the entire study design and sample size estimation, and maybe even into what type of design we actually look at. So the amount of discordance really is inherent in the study design, and we don't need really a threshold unless we get through part of the way through the study and we find out that the discordance is much greater, assuming no bias. All of this assumes no bias. Once you throw bias in there, all bets are off, but if the discordance becomes greater and there is no bias, we don't have an adjudicator or the terminal read is not biased and the decision is made not on a biased read, then all of this is inherent in the study design and should have been with the statistician involved from the very beginning.
So, with the statistician involved in the beginning, the study design has these issues inserted into the study design, into the sample size estimation, and maybe into the randomization, hopefully the randomization.
What we can do is we can figure out that the terminal read is the answer and then go with that, but I have colleagues over here that would strongly disagree with that because they have seen the results that the adjudicator or the terminal read can be strongly biased, and that may be when we do a review of the read, look at the discordance analysis.
So it is a fairly complex issue. It is not an issue that I think we can put a 30 percent or 20 percent on, especially if the study design has inherent into that, into the sample size and into the study design, the issue of 30 percent or 20-percent discordance. If we have 30-percent discordance and we have an adequate study design, then there is no reason to put a threshold on at 30 percent or maybe even 40 percent.
The other issue is that I think -- I am not sure if we have addressed this before, but discordance also can mean reader drift or it can mean reader training, and that may be why we have to look at this, to determine whether we have reader drift or reader training that is required, to get them back onto a level playing field, so that we are not confounding results with time, reader drift over time.
Actually, I would like to hear what the FDA has to say about the terminal read and using them as the final answer, rather than a consensus, best two out of three, because the terminal read really is -- if you look at it statistically and talking statistically, it is using prior knowledge from the first two reads to update a final probability that you have a diagnosis or that you have your endpoint.
This gets back to what Wen-Lin was saying, and that is define the endpoint and then go after that.
Thanks.

DR. LUO: I think there is one more point I want to make. The discordance issue probably will be more important in the endpoint based on the absolute magnitude, like the oncology. You would define progressive or stable. So that probably, the reader may have a different conclusion.
However, in some endpoint, that is the change on baseline. If we have the patient's image all reviewed by the same reader, then the bias could be minimized if we use the same reader to read all the images from that patient, but the magnitude, yes, probably the discordance issue will be higher, but we can use some kind of blinding reading or the randomization, that kind of thing.

So we can use other ways to better design the whole study to minimize those biases, and as David said, the training for the reader is also an important factor in the study. So we want to use the well-trained reader in our study.
DR. MARZELLA: I just wanted to add another comment which is also the issue also for the agency is one of generalizability, and I know that it is more applicable to specific therapeutic, specific types of incidents, but the concern is if there is a treatment effect or diagnostic effect that is only seen in a specific content and it is not broadly applicable, that would be very concerning.

This is particularly the point, for instance, for diagnostic agents. At the very minimum, it would suggest that a training program is necessary to make sure the diagnostic agent is being used in a manner that produces useful information when it is used by clinicians out there. So there is also, again, the issue of generalizability.
However, let me step back and make a comment about how the field is moving. It is so delightful to see statisticians being involved in this discussion, and I would like to echo what has been said from the podium, that the earlier you can get the statistician involved in understanding what these issues are, the better off you are going to be.

The analogy I would like to give -- I have been doing this for many years -- is at one point, it was not uncommon for sponsors to send in clinical protocol that did not include the statistical analysis plan, and the idea was we will wait until the study is completed, and before unblinding, we will send in the statistical analysis plan. Well, it is a recipe for disaster.

Here again, the idea that I think is going to be helpful is to have this level of integration from the very beginning where operationally you have a statistician, again, being acquainted with the issues that can come up in terms of where is the variability that I need to worry about, what level of perspecification needs to be involved, so that we have the same concept.

For instance, in therapeutics, therapeutic drugs is the concept of intent to treat. It is very important. Well, the

concept of intent to diagnose is also very important. So, if you don't have a standardized predictive way in which you are going to acquire images, select images, determine what to do when you have missing data, all of these issues really need a lot of statistical input from the very beginning to make sure that the data is of high enough quality and is interpretable.

So I am delighted that this discussion is taking place. I think it is going to be very important in facilitating the success of trials that use imaging for endpoints.

DR. MILLER: I think you have raised some fascinating questions and thoughts, drilling down almost to the fundamental question of why do a central review at all, and that is essentially where I think one of your questions was going.

Picking up on a theme of several of the other folks, including Dr. Marzella, I think there is a couple of fundamentals there. We are talking the difference between patient management and clinical trial evaluation where we are trying to get a quantitative endpoint in a very stylized manner.

As has just been pointed out, it is everything from how many images have you got, how that is being presented, how the QC is going. All those aspects are going to make a change to the way the central review is conducted vis-a-vis the local site review, which is going to be conducted in a totally different format.

Probably, the major difference there, particularly as the trials get larger and larger, is many of the radiologists at the local site haven't experienced clinical trial evaluation, have not necessarily been through how to determine, in this case, resist criteria or WHO criteria.

In other therapeutic areas -- and let's broaden this out for a moment -- you go for rheumatology, go for osteoporosis. The radiologists that are undertaking those analyses, although those interpretations are very specialized, we heard yesterday that in the rheumatology field, there is probably only about a handful, maybe a dozen folks that routinely do the radiological evaluation for RA. Osteoporosis is about the same. There is about 20 radiologists in the world that undertake that.

I would also then entertain that in the oncology field, there is a group of specialist radiologists that have either gone through training or have had experience that is, therefore, going to encapsulate the radiological evaluation a little differently from the local site evaluation, where there is going to be a lot more variability and even understanding in how those reads should be conducted.

I throw that out. Therefore, I would expect there to be a totally different set of data coming from the two. The central read, in my view, is going to be a much more concise, very determined, very focused read compared to a very broad spread of evaluation, and I think with that understanding, what you are getting at is very definitively two sets of data giving you two totally different stories, information that is useful, but very specifically, a central read is giving you the radiological evaluation by a very controlled set of QC, how that data has been managed throughout the whole process, very stylized in who is reading it and how it is approached, and I think that is where a lot of the information and the change is occurring.

I throw it out as a further discussion point.
DR. SRIDHARA: We really don't see a huge difference between investigators and independent reviewers where we have accepted the reads, and we have been able to explain why there is a difference if there is a difference. As I said, in certain situations, it is that the target lesion itself is different. So they are really looking at two different pictures, and therefore, you are getting different information.
True, we can just deal with using only the central reading, and that is what we say generally in open-label studies. We prefer the central reading as the primary way of assessing the progression.

Having said that, the problem is because there is going to be a change in therapy, the questions come when the investigator on site has declared a progression and changed the therapy, and the independent reviewer doesn't get to read this for quite sometime and reads this as a non-progression. Then what do you do with it? Because an action has already been taken.
So it is more to deal with how quickly or how much in real time the independent is able to read these versus -- we have also seen in one particular case, you have a data cutoff, and at least some of these can be minimized. You have the data cutoff for your analysis. We would expect that all these scans are read up to that date by the independent reviewer. We have seen that that doesn't happen. It is the last batch until that date, whatever the independent review has seen.
So there is really totally different information that you are looking at from investigator and independent review. That can be minimized, and at the data cutoff, whatever the site has reviewed, all of it should have been reviewed by the independent review as well before we make some conclusions.
As I said, there is a thought whether at randomization you mark the lesions and the same lesions are tracked by both groups, so you are looking at similar things. These are some of the things that can happen that could realize that.
Also, the protocol itself is written, so that every investigator follows the same thing, whether it be delivering the therapy itself or whatnot or who should be the patients that are to be enrolled. Similarly, we can have for the site readers as well that maybe an education or how the read should be or what is the criteria for progression and so on and so forth.
DR. ZALKIKAR: I just wanted to make a comment. It has been bothering me. You were saying what if the on-site readings and the off-site readings lead us to the same conclusion, but there is this huge amount of discordance between them still, somewhat unexpectedly high amount of discordance between them.
I can sort of put a scenario in front of you that the on-site readings are under the open-label setting. So we have concern that maybe bias is being introduced.
Now, the off-site readings, when we look at the discordance and as to why the discordance is occurring, maybe the off-site readers didn't get to read the images that the on-site readers looked at. They didn't get passed on to them, and there were imputations going on for the off-site readings. Then the imputations are driving the

statistical significance there, and the on-site, maybe the bias is being introduced. We don't know how to delineate that bias or take that apart statistically.
So how do we proceed now? The whole question, the measures that you were saying to put in place or to assure us that the study has been conducted properly, a lot of these things, if the study has been conducted properly, the study conduct issues are not there in the picture. A lot of the discordance could be minimized to a substantial extent, and then whatever is left over is sort of inherent in the process.
DR. MILLER: I would like to follow up on a couple of those points because it kind of adds an interesting question.

The first one was that the local site may determine it progressional or not. The central read will make a different evaluation. They are not being read contemporaneously. I think you are absolutely right. That is going to happen because you have got two different things going on.
The local site is managing the patient with a lot more information to play, which is very appropriate. That is exactly what the local physician should be doing is managing the patient.
The central read is not providing patient management in any shape or format, and therefore, yes, everything should be read, but there may be disruption and a difference of opinion, but it is a radiological interpretation, not a patient management interpretation. So that creates that difference there.
The other question -- and this goes back to historic acceptance -- central reads across all therapeutic areas has been that the images are stripped of all markings from the local site, and the interpretation is devoid of any prior evaluation from the site.
You are suggesting that the local site now marks up the lesions, and these are the ones that are carried forward. That is a total difference to where everything has gone historically, and I would just like a further comment on that and how you would elaborate on that, please.
DR. SRIDHARA: I don't think that is the best method. I am just saying that is one method where perhaps you are seeing the same scans then. At least you are not seeing different lesions, but that is not what is being used, and we are not right now recommending that everybody should do this because the whole idea of this independent review has been that it should be a totally random selection of lesions rather than somebody directing them that these are the lesions to be picked.
So that kind of blinding is gone when you choose, when you market and say these are the ones that you have to follow. Certainly, there is a problem even there as well.
DR. DORFMAN: Just a couple observations. Yesterday, we spent a lot of time talking about how to make sure the imaging review is truly independent. So the folks at the site cannot be independent reviews. They can't review lesions form their own, supposedly to elevate the quality of this imaging review.
To use an analogy, let's say, from the lab world, it would sort of be as if you had sites where everybody mixed up their own lab kits to do a particular lab test, and then centrally, you had a qualified machine, let's say, that did this

in a very quantified way. You say, "Well, but there might be a discordance," and you would say, "Well, of course, there is. These people are all mixing up their own stuff on the bench top, and we would never accept that." The discordance is they are doing things in a very, if you will, cottage industry way, and we are bringing everything centrally and doing it in a very scientific quantified way.

Well, we are not up to fully quantified imaging yet, but this is, if you will, more than a baby step to kind of get there and to move things to an independent review.

So, in terms of the issue of site reads that lead to changes in management, because as was pointed out they are taking care of a patient and there are management decisions there, a patient can come off a protocol for any number of reasons. It could be a clinical decision by the caring physician. It could be because of comorbidities.

It could be because they got hit by a bus, or it could be because of an interpretation of an image. Either way, they are done with the protocol at that point.

So I think it is important that whoever is -- and this is way off scope, David -- for whoever is designing the study, to design the study to account for a certain amount of dropout for a variety of reasons, among them a local read of an image that might contribute to a decision that takes the patient off the protocol.

Designing your study to account for that is really important, obviously, so you have enough power at the end, but at the end, when you have people who actually didn't come off the protocol and you have discrepancies, what you are talking about is a truly independent read done in a scientific manner as part of the conduct of a trial versus a read that was done in a local manner as part of the conduct of clinical care.

Now, there are ways to try and square those things out. For example, you can recruit all the local readers to read a test set along with the central readers and determine if there is discordance related to their skill level, or you can say that the fact is they had different software tools to do lesion measurement at the sites than they did centrally, whatever it might be, but you start to add layers and layers of the onion here which are probably unnecessary if you just make the point that the central read was designed to eliminate all those variables. Then trying to describe in a scientific way why doing it right is different than doing it as part of clinical care, it seem contradictory to the whole purpose of what this is about.

Maybe I am just missing it here, but I am curious as to why this is an issue.

Clinical care, you need to power your study to make sure that clinical care doesn't unduly influence your study and undo all of your design. That is part of study design, but once you do your study, why is it that the clinical care data that are determined completely different than the conduct of science are suddenly entering the picture, and why is this analysis important?

I apologize for being thick, but I am just trying to understand that.

DR. SRIDHARA: I think you are absolutely right that for our evaluation, we will use the central review data and probably not the on-site data at all, knowing that is how the protocol was designed and that is what we will be

using.
I would go to further extend that it would be a problem even with just off site if there are two independent readers and there is a discordance between the two of them.
So it is not just the discordance between off site and on site, but whatever readers, even off site, whether there was a discordance and how do we deal with it, that is an issue as well.
DR. DORFMAN: Agreed. So clearly, the scientific plan needs to deal with how you deal with discordance of the central read and adjudication and all that, but would an acceptable policy in terms of eliminating the discordance between central read and clinical read be to just not have the clinical reader fill out a case report form? I mean, that is the clinical care of the patient. Why collect a CRF from the site at all with regard to the image?
If those data are not part of the scientific conduct of the trial, why are those data important to collect?

DR. SRIDHARA: Because basically, as you said, it could be that the patient is going off protocol or that they are getting some other therapy or if there was some other information or if the on-site reader had actually different information and looking at it, we want to look at it and see whether something was missed by the independent reviewer as well.
So, even though, yes, it was predesigned that way, we want to make sure in the end that the data that we have is the true data and it is a reliable data. So, in that process, we need to look at the on-site read as well.
It is not that that will be the information that may be used in the regulatory decision. However, we need that.
Let's say you see there is more censoring going on in one arm versus the other with either of them. Then that is a concern as well.
DR. DORFMAN: Well, there should be a censoring form or leaving protocol form, and if that is secondary imaging, that ought to be captured.
DR. MARZELLA: Let me just take another turn at it. The issue is what is the importance of collecting this information, and the importance is that it speaks to data quality. So the issue, again, is not that there is going to be -- we are all agreeing. I think everybody is in agreement that the answer that you get from the blinded read is designed to be more reproducible, but we want to also know how reproducible that difference in quality between the two is because sometimes it matters.
Also, the issue then is the issue of generalizeability. The issue is in the real world, is this diagnostic working, is this therapy working, or is it only such that you have to have a highly artificial environment where you can show that it actually works. So it is more than an actual experiment.
We are also asking in actual clinical use, do we get the same answer. Ideally, the thing should be trending in the same direction. It should be statistically significant in the blinded read because it is less variable. You have more power. So you have more control, but if you get a completely different answer, it works in one setting and not in the other, then the issue becomes how reliable is the blinded read to begin with.
So the issue is very important. You need to take a look at all of these. We are not saying it is another hurdle you

have to me. It is another thing that is important to look at for data integrity and to figure out how generalizable this effect is to have more confidence, that, in fact, it is going to translate to actual clinical use.

DR. UPRICHARD: So my comment is along the same lines. The April 2005 guidelines that were put forth by the agency on the evaluation of oncology agents really specifies that if you are going to look at progression as an endpoint, that you use a central imaging facility, and that is kind of the final say.

I am fine with that. However, there are some instances such as CNS tumors or pancreatic cancer or even lung cancer where I have had investigators say, "Look, the CT scan hasn't changed. It shows stable disease. That goes to the central imaging facility." However, the patient is having more seizures or they have clinical progression, such that they take them off the study. They refuse to check the box on the case report form that says this patient has stable disease and has progressive disease.

So that actually then increases the opportunity for discordance, and that the central imaging facility will see stable disease based solely on the CT scan and will not have the benefit of the data that suggests the patient progressed clinically, which probably is relevant in terms of time to progression.

So I would like to see, as this guideline is developed in 2011, some statement to that effect or some allowance for clinical information or clinical progression information be provided to the central imaging facility.

DR. MOZLEY: Can I make just a quick point? Correct me if I am wrong, and the rest of my colleagues, too. We collect the clinical data for the benefit of the data safety monitoring board that is going to do futility analyses for safety and toxicity.

So the key metric there is safety and toxicity. A lot of patients go off a study because of dose-limiting toxicity or intolerable side effects. Even in oncology, it is not necessarily because of disease progression.

The key point is that from our perspective, the efficacy readings by the sites are not relevant because they are not conducted in a rigorous scientific manner by scientists whom we have had the opportunity to train, and when we are running a 150-site trial for a total of 300 to literally 600 patients and there is only two to four cases per site on average, sometimes only one, we are not particularly confident in the outcome measure generated by the local readers.

Again, as it has been pointed out, they are two different datasets. They are two different outcome measures. They are two different drills. They are not comparable, but the safety data, nevertheless, is more authentic, if you will, in the hands of the local site investigators. Whereas, the efficacy outcome measure is better established by the central readers.

So we need both pieces of the puzzle, but the comparison in the efficacy outcome measure between the local site readers and the central readers is not of particular interest, as I understand it, and push back on that if I misstated the case.

DR. LUO: I think that point is -- unless in Merck, it is true that the efficacy only based on what you read from the image, but if there is any seizure or any other side effect or any accident happened to that patient, that is out of

story.
Clinical monitor always monitor those patients day by day, time by time, and then if anything happened, they can stop the patient in the trial at any time, but that would not be the criteria to make the decision about the disease status or the endpoint.
DR. KRESS: In many of these studies, not only is there a radiological or image review by primary reviewers and then an adjudication, but then there is an additional step whereby the clinician, the oncologist sits down either with the radiologist or independently, reviews the radiological data and prespecified uniform clinical data and then makes the ultimate decision on the time point of progression of disease. So that incorporates the clinical information at that point in time, and that is incorporated in many of these studies.
DR. TOTTERMAN: Actually, I had a question regarding the reading, the on-site reading and the off-site reading. When the disease is progressing based on the on-site reading, the patient is taken out from the trial. This is considered as the final truth from that regarding what the off-site readers are saying.
So then I had a concern regarding that, that maybe that wasn't the correct reading after all, but the patient is off the trial in any case.
My concern actually has been I have been on both sides on the fence. I have been doing in a clinical environment, as well as doing a central reading, and I did the clinical environment for a very, very long time.
I echo Colin's concerns. The clinical readers usually are not trained. It is very few occasions when they actually train exactly for this clinical trial and train in just a way that they actually know exactly what they are supposed to do, what is the selection criteria for the tumors that they are supposed to measure and what is the target lesion in general at all and how they are supposed to measure.
Regardless what tools we were using and they are using, all the measurements are not always that high, and the variability of the measurements is also not that great either.
We have done these studies several times. We are measuring the so-called known trained as well as trained readers measure -- or evaluating the accuracy or variability of the measurements. The trained reader who is actually trained to do that, who goes through a rigorous training, the reproducibility of the measurements or evaluation is pretty high. So that is my concern.
One other concern is also that when I am the reader, off-site reader, I see cases where the disease is really progressing, and I see time after time after time them coming through. There is very clear progression of the disease. So then I have the concern about that, that if the disease is progressing and we are evaluating it in any case, the patient is on trial, but the patient who maybe I didn't see they are progressing were taken off the trial, and they were never followed with the really true course.
Of course, there will always be the clinical data or the clinical evaluation, where the progression actually happened, but if there is no clinical, then the question is what is the real truth in those cases where the patient was taken off the trial.
I guess the comment is, is the resist really the accurate measure, the patient measurement to evaluate tumor

progression in any case, are we using something less than perfect measurement to evaluate the disease progression.

DR. SRIDHARA: I will see if I can try to answer some of it.

First of all, I think the whole problem is that this progression -- again, I am speaking on oncology -- is a very subjective assessment in whichever way you look at it. So it is a less than perfect measurement for evaluating any therapy, and therefore, the first choice should be survival where there is no discrepancy there and nobody can argue with that.

But having said that, it is true that if we do believe that the disease model that everybody has to progress before they die when they have cancer, then it is reasonable to pick time to progression as an earlier endpoint.

I am not using the surrogate endpoint here. I am not saying that it is a surrogate endpoint, but for a patient, progression itself could be a benefit. There is lack of progression itself could be a benefit in itself.

So we are not talking of surrogate endpoints, but just the progression itself is a very difficult endpoint. It is a subjective endpoint. It is not only dependent on the technical ability of the reader, as well as what instruments you are using or what methodology you are using, but it is also dependent on assessment timings, how frequently you measure or how much apart you measure, and also, for whatever reason, you take the scan and it is not readable. So you suddenly have a missing piece there. So these are all issues with progression data.

It is not just if we can move away from the discordance/concordance part of it. That is an issue, and we have to look at it and see whether we can explain some of it or we cannot explain some of it at all.

The bigger issue is how do you define this time to progression or what do you call as the progression date, and yes, the blinded review should be used, but let's say the blinded reviewer felt that there was progression and the on-site reviewer said there was no progression or vice versa. There is a consequence to that.

So, if there is a change in course, then this time to progression evaluation changes. You have to sensor, do something for that data. So there is no way not to collect the site data. It is very important, and we need it, and if there is a change in course that is needed as an information to see where do you put this date.

As I said, the whole thing is because it is not assessed every day, that you don't have the continuous information, and it is very much dependent on the frequency with which you are taking these scans.

Maybe the independent reviewer is reviewing all the scans at regularly scheduled visits, and here the patient may come with some symptom, and so there is an unscheduled visit as well. So the on-site reviewer may be looking at statement different which was not a scheduled visit. All these things play into the assessment of time to progression itself, and therefore, we need all of the data.

DR. CONNOR: I have a concern about the general use of discordance then in this case. Discordance is the result of two different answers using the same dataset in the same process. The sites don't have the process, and the central reviewer doesn't have all of the data. So how can you establish discordance when you don't have all of the same data and run it through the same process?

DR. SRIDHARA: I think that is what I was trying to get at, that, yes, you cannot truly address this, and the whole

idea is can you explain.
If you don't want to call it discordance or concordance, you can talk about is there a difference in the dates of what you call progression.
DR. CONNOR: That would be the way to approach it then --
DR. SRIDHARA: Yes.
DR. CONNOR: -- and abandon the term "discordance" because it does evoke a very strong reaction which then has to be rationalized.
DR. SRIDHARA: I agree. That is why I was saying how can you talk about it if they are not even looking at the same target lesion. So you have different information at both sites, and so you expect that there is a difference, but our concern is where do you pick that date of progression, and which one do you use, and we will have to make use of all of the data that is available.
In a particular case, if there was overwhelming clinical data to suggest that there was a progression, even if the blinded reviewer did not pick up that date or they didn't have the information and that is why they didn't pick it up, we will have to include that in our analysis.
So it is not a very simple straightforward analysis, but it is the overall information that we have.
DR. FORD: I was wondering if you thought that simple reconciliation of site versus central discrepancies prior to submission would be something that would give you more confidence in the data and would eliminate this issue. That is the first question.
The second question is there are other models of use of central data elsewhere in this industry, the use of central EKG data, and I realize that is for safety. There is also the use of central hematology data, and I realize that might be for eligibility review, but I think in those two instances, there is little reconciliation of the site versus the central data. I just wanted to point out that perhaps in this instance, your model is a little bit different.
DR. SRIDHARA: What we have heard is we would love to see if somehow a procedure can be returned in how this can be reconciled, and sometimes it may not be possible at all, or at least have some rule of where you see the difference and which ones can you reconcile and where you cannot.
The problem that we have heard is that the expert radiologists are not available all the time, and so they are hired for specific time periods, where in batches they get all the scans to read. Therefore, this kind of different information looking at it and all of this is coming in the picture.
If it is happening in real time, as I said, the EKG or some of these, probably the data is done. It is sent immediately, and you are to read back, or when it is a part of an eligibility criteria, you are looking to enter the patient. So you want that information right away. So you are getting that read back immediately.
In this case, it appears it is by design not on a real-time basis, but supposed to be happening periodically.
DR. FORD: To expand on that, to ask my question, irrespective of those differences, is that something you would find helpful?
DR. SRIDHARA: There have been some where we do look at it and see whether we can reconcile. As I said, if

there is overbearing some clinical data, that the blinded reviewer did not have and that is why the site had called it progression, then, yes, we do reconcile within our agency since we have data from both datasets, but it would be good to have a procedure up front set up on that aspect.

DR. NOURSALEHI: So in a situation where your way of collecting progression, for example, in your case where you are talking about oncology, where the standard of care is less than adequate or less than perfect and you know there is going to be discrepancies and discordance, I think everything we said here maybe has to be looked at from a different perspective.

In other words, you should not be looking at or maybe we should not be looking at everything in the same light in terms of adjudication rate. So, yes, it is therapeutic area-dependent. Yes, it is sensitivity or specificity-dependent. At the same time, it is what kind of options do you have available for standard of care in a certain therapeutic area.

So, if your choices are limited, then you have no choice but to live with much higher adjudication rate. That is one issue.

The other thing is when you are in a situation with, let's say, off-site readers, your reader one, your reader two, your adjudication, your worst-case scenario, your investigator, even though there are differences amongst them, but they all point in the same direction and demonstrate the same superiority of one arm over the other, wouldn't that be enough? Particularly in light of the fact that each of these readers, despite their differences in readings, they look at both arms at each site that they are responsible for.

So whatever biases are brought into the process, it is balanced across the sites or balanced across the patients or balanced across the treatment arms. That is the other question I have for you.

But then also I have one more question, and that is survival as the gold standard in oncology, for example. I am not quite sure if that is the case. That is up for debate because in situations where patients live long and they go off treatment because they are assessed for progression, your placebo patients are exposed to less toxicity, and they had a drug holiday. So they go on, off the study, to other treatments. They have now added advantage to live longer.

If it was ethical for us to control these patients under the same umbrella as the study, regardless of whether they had progression, to continue on the same treatment arm, then, yes, your survival data would be pure, but in cases where your patients go off the study onto other treatments, I am not sure if your survival data is your gold standard anymore.

I would appreciate some comments on that.

DR. SRIDHARA: I will start with the last one first.

If you ask a patient no matter what sequence of therapy you take, a cancer patient, what counts is how long you live. It doesn't matter when you progressed with what or what was the sequence of therapies that you took.

In the overall picture for a patient, the benefit is to improve survival. So there is no question that the whole development of oncology drugs, the aim is to improve that survival.

You give this and you give the next one something else, but in the end, it is how long you lived, and there is also a philosophy that if in the early time on, if you have had stable disease for a long time and your time to progression is longer, then that should affect your survival as well.
Yes, it is confounding, and it may not be the best way, and it may not be as randomized anymore or whatnot.
So, from a patient perspective and clinical benefit point of view, survival is a gold standard.
DR. NOURSALEHI: But then your placebo patients most likely are subject to less cycles of treatment during the study and also less toxicity. How would you explain that?
So they go on and they get off the study sooner onto other treatment. How would you account for all those differences? Obviously, survival is confounded.
DR. SRIDHARA: It is confounded, but in the end, it is important. Placebo patients, if they are going to another treatment sooner, it means that they have progressed further in their disease as well. So they are worse off already in their disease.
DR. MOZLEY: Many of your points are well taken.
I would just like to point out that the field is rapidly evolving under our feet. When you have highly efficacious drugs, then it is unethical to leave patients on trial who are continuing to progress.

Those progressors have to be crossed over, and in those situations, you will not have a difference in overall survival, only progression-free survival.
So it is one instance, a happy scenario, where dogma about overall survival being the clinical gold standard may be eroding as we speak, and it brings up the point that we are not always going to have highly efficacious drugs where we can cross people over rapidly, but it means as a field, we have to develop and validate progression-free survival, so that we can use it as a tool more effectively in all scenarios.
DR. SRIDHARA: I think I made the point that if we assume that the disease model is such that progression occurs prior to death, then using time to progression or progression-free survival, either of those two, seems like a reasonable endpoint to use.
The problem that we are facing is in the measurement of progression itself. Here, we are discussing the same point of which date do you pick, and so statistics can help only if you know the data is correct and true and is the same, whoever looks at it, whichever way you look at it, it is the same date.
So here, it is a subjective endpoint. It is dependent on how frequently you measure. If your frequencies are once six months apart, I am taking an extreme. I know that probably is not the case. And you didn't have progression at this point, but six months later, you measure and there is a progression that you detect, probably it happened somewhere in between, and it could have been closer to this end rather than that end, and we don't know that.
Given that situation, it is always the problem of how do you deal with these issues, and when the clinical site is doing something else, what do you do.

DR. NOURSALEHI: In the situation that you just explained, if your reader one, reader two, adjudicated, the worst case is what you just explained, worst case meaning the earliest progression date for your active arm and the latest progression date for your placebo arm. If all of that, including the investigator, all point to the same direction, the same results, the same conclusions, would that still be a concern?

DR. SRIDHARA: I think it will have to depend on disease setting, and I don't think I can answer that question.

DR. SULEIMAN: I think the adjudication process is interesting, but I think you are missing some real basics.

I think if the imaging task is more clearly defined and if the readers are more carefully trained and consistent with each other, I suspect the adjudication rates would drop a lot, and so I think you should focus on the basic fundamentals more in terms of what is it you are looking for in an imaging trial as an endpoint. I think the adjudication process is a real wasteful process because it should never be dominating the conversation. It means you have got a problem further down on the chain.

I think some of these imaging tasks are poorly defined, and it is obviously, as somebody said yesterday, a lot of the observers are not necessarily trained or periodically trained. So I think if you address the fundamentals, this issue may get much, much smaller.

MR. GASTINEAU: Well, with that said, we might move to the next slide.

MR. GASTINEAU: It is pretty close. However, I do think many of the topics, perhaps we have already touched on as we kind of moved forward.

I may take the liberty of kind of skimming through a couple of these things to see if we have already touched on them and to see if there are salient points that still are resonating within the audience and within the panel here. Maybe we can come back to this in terms of the guidance that we hope to put together. It would seem almost a foregone conclusion that what needs to be stated within the statistical plan, as well as the IRC, is which of the datasets are going to be the primary dataset and define that up front.

God, here we go with phantoms again.

MR. GASTINEAU: So that one is on a point of consideration.

Modifications to criteria, it is usually included in the protocol. So, again, the question from our perspective was how much content do we want to put into the charter. That is where we started this discussion, and what is part of the SAP as well.

Reader training is an issue that I think we all addressed in very different ways, but we clearly recognize reader training as a hugely valuable endeavor in terms of trying to reduce variability and get people on the same page. In fact, as someone has already pointed out, the main reason for doing independent reads is the reduction of bias and a more reproducible process, which we believe is inherent in this particular process.

Reader training, as I think one of our statisticians pointed out, really is an underpinning of the refinement of design power, sample size, and those type of things. If we understand where our variability is coming from, then they can power accordingly for that.

Redesign, missing data, imputations, those type of things probably are worth coming back to, as we have got a

few more minutes.

The question that has been floated around is the default, I guess, at least in the current standards around some of the oncology and other things, has been where the reader ultimately may be working with an algorithm, maybe a criteria scoring system, whatever those things are, but ultimately, that medical opinion is the key outcome measure.

So, while those criteria are part and parcel of their outcome measurement, we have let the reviewer essentially give us the overall outcome based on that information. The question that has arisen is, is that the right model, or should we let the statistical models with imputation kind of go forward to guide those processes.

Adjudication, as we talked about or heard yesterday, was not really considered a measurement of quality, but really a function of the design and the systems in which are in place, and the annotation that readers have different opinions, not necessarily wrong, different views of the data and styles associated with that.

I think for part of this, the definition within the statistical plan and within the charter that was part of the discussion was what triggers adjudication rates, what are the things that we will look at in terms of comparison.

In terms of talking about adjudication rates, defining up front how rates will be calculated, I think or those of you who saw Dr. Schwartz's presentation, there clearly are differences in what we would describe as a numerator and denominator and categorical responses such as stable disease or partial response versus a data progression where we have a continuous variable that goes across many, many different scans or time points.

So those are the type of things that I think are critical in terms of understanding a response rate or an adjudication rate.

Blinding and randomization were on our list. Randomization was talked about a little bit yesterday within Working Group 2 in terms of how the images are presented, sequential views, sequential unblinding, randomized, those type of things, again, points of discussion within the statistical group, as well as blinding, and the notion of complete blinding from a pure radiology perspective, the additional clinical information and sort of a sequential unblinding piece and what do you define as the ultimate outcome for those type of processes.

Inter- and intra-reader metrics. Again, this is essentially the idea around reader performance, reproducibility. The question really comes up of what do you do with this data.

The conversation within our group has been really the way we have traditionally looked at inter- and intra-reader reliability has not been powered for statistical measurement, and so it really has been a tool by which we simply survey readers' exercises, so that we look at training exercises or retraining of readers or things to kind of keep them within a performance metric that we are comfortable with, but not to compare inter- and intra-reader reliability at a statistical level.

Finally, data integrity, which was part of our charter as well, I think inherent in a lot of the things that we are talking about and the processes that are involved in this, the collection of data, the audit trails that are associated with that, the off-site and independent processes, reduction of bias, inherently all go back to the data integrity question.

So that is sort of what we have got, and probably another 15 minutes or so, Constance?
So, within that context, should we run and hide, or are there points that are more on people's minds than others? As we have got our FDA colleagues here, are there things you would like specifically addressed?
Tom?
DR. FUERST: Yes. I would like to talk about reader agreement, but maybe not in the context of oncology trials. This question is about whether a single reader should read all the cases for a study or from a pool of readers, issues of replacement of readers, and from a statistical standpoint, questions about measuring reader performance and making a decision about whether Reader A can replace Reader B partway through a study or we can use two readers, each reading a subset of the whole population, and the types of tests for reader agreement or the expectations of reader agreement that we might see from the agency in the context of other types of measurements, whether that is a quantitative measurement of a volume or of a hippocampal volume or some other feature of the brain or another part of the body, cartilage volume, and then in the context of other subjective readings, like the sharp scoring of erosive and joint narrowing changes in rheumatoid arthritis, and then related to that is something that you propose in your document, the question of coefficient variation of these measurements, is that something that the agency is interested to know on a study-by-study basis or a technology basis, or any thoughts about that which I think are relevant statistical points.
DR. STEVENS: I can address that.
One of the things that we need to keep in mind -- and I was reminded of this yesterday by a couple of the radiologists in here who have memories that don't ever forget. Many years ago, I gave a presentation on inter- and intra-reader variabilities and assessments thereof.
One of the things that got the radiologists' attention was that I called them "nuisance parameters." They didn't take too kindly to that, but the thing it reminded me of was I talked about doing 50 readers to assess all these measurements that we are talking about here, and the question is would you run a clinical trial with two observations, two patients, and the answer is, "Well, I would love to, but the agency wouldn't let me get by with it."
So why would you want to measure inter-reader variability with two readers? You have two. What kind of information are you going to get from that? Not very much. So you would need 50, 60 readers to get a good assessment of true inter-reader variability or inter-reader agreement or whatever you want to call it.
So, from that regards, the whole concept of setting up metrics for inter-reader variability, reader-to-reader variability, et cetera, is really a moot point. We calculate these measurements, but what do they really mean, and how can we interpret them? Because we really don't have good data on inter-reader variability. So I think that takes care of a whole lot of these things.
I really like your points here, though, especially Point No. 5, which is what I was just talking about. We calculate these things, but they are not powered for anything. So either we calculate them just because they are there and we want to look at them and that is it, or we don't even bother calculating them at all and just get on with the

important stuff which is taking a look at the blinded reads.
DR. RAUNIG: The importance of reader to reader, inter- and intra-reader variability is typically -- well, it should be used in the design of the study, but in all cases, the reader protocol really must be considered in the statistical design and for the reason that you pointed out. If you have one reader, at one extreme, if you have one reader and you come up with a conclusion, your inference is to that reader and that reader alone.
So, if you want to expand your inference to all readers, you have to have a good sample of the population of readers. Whether that means you have site reading and you have off-site reading, on site and off-site reading, or whether you have a central read and you have 50 readers from a central read, in that case the statistical design, the study design has to include that in the randomization scheme and the powering scheme and the variance analysis scheme to be able to come up with a design that is going to give you an answer that you want and with the inference that you want for the study.
If you come up with an inference that applies to three readers and you think three readers is a good sample of the population, then you can say that, okay, well, this drug is going to work for all readers, this drug is going to work, this drug is globally efficacious, but if you had one reader, two readers, three readers and you don't think that is a good sample of the population and you need more, then you need to consider that in the study design.
This goes to other design issues, too. It is not just this slide, but it is the other slide, too. Other design issues must consider what the inference, what is the answer you want to get. If the answer you want to get is does this drug work if I use central CRO A, then that is not going to fly with the FDA. I doubt that you would approve a drug if it only worked with one central reader, and you couldn't expand that inference to all.
So the design issues really have to deal with readers. They have to deal with this intra- and inter-reader variability because that is going to be part of the study design. It is going to be part of the overall sample size estimation.
Other topics, the other topics that we have brought up here, all of these issues, including endpoint, which is one of Wen-Lin's major areas of concern, it is to define the endpoint and then define the study and then justify it and stand behind it.
So, if your endpoint is you got some disease or stable disease or complete response, then define that endpoint and then stand behind it and run that test to do that and define your inference, stand behind your inference, and run a test and design a test that actually you can estimate whatever estimate you come up with draws a conclusion on that inference.
DR. LUO: I think to answer the gentleman's question about the replacement of the reader, I think, yes, you can read it if something happened unfortunately to that reader, and you have to, but the question is -- or the point there is that the same reader has to read all the images for the same patient, instead of first half by one reader and the second half by the other reader, because probably that will cause a different baseline. This reader may say this size is 50, but the other reader may say that is 100. Then that will cause the bias.
So we want to focus and emphasize that the same reader has to read all the image data for all the patients, and

also, this reader needs to be consistent. So that if there is bias, then there is the same bias all the way through. At the end when you do the calculation that could change baseline, then that bias will be kind of the cancel-out.

DR. FUERST: I agree with those comments in a lot of cases, but mostly where the reading is a very subjective one, and there certainly are measurements that are more quantitative, but still might be subject to the decision-making by the operator. If one documents that Operator A and Operator B measures in volume similarly and a reasonable number of test cases adequately powered statistically, do you then have the same requirement that that same reader measure the volume at very visit for that patient? I think one can begin to make an argument that it doesn't matter anymore whether it is the same reader or different reader. You can produce evidence that it is not really a requirement.

So I think we have to think about what kind of a valuation tool we are using, the degree of subjectivity, is it more quantitative or more subjective, before we can make some of those decisions, and what you said, I agree with in certain circumstances, but in other cases, I would think that it is not necessary to have a single reader read every case.

DR. LUO: Right. I think under the condition that you will control and will design your study, then probably you can say no, but in case anything happened in your procedure or actually some unbalance of the study or whatever, then it is still preferred that same reader for the same subject because we don't know what would happen in the study.

For the question about assessment of the inter-, intra-reader variability, of course, you can do that by probably early kind of the pilot study to assess that. However, you can assess that for certain group of the reader, but can you really apply that to the kind of global reader? And different readers probably have different variabilities.

So I am not sure how much information that can provide you, and rather, I would prefer to focus on the variability of your endpoint, and actually, that is the variability that we used to power the study, to calculate the sample size. But it is true that we need to know all possible resource of the variability in our data, so that we can better estimate the overall variability to power the study.

DR. ZALKIKAR: I just want to make a brief comment about the term "diagnostic imaging" point of view.

In diagnostic imaging, I would think that intra-reader variability, whether the reader is reading the same thing again and again or not, is important, that assessment is important, but that really needs to be made ahead of time, so the reader training could be put in place rapidly in order to address that.

The inter-reader variability in diagnostic imaging, in diagnostic imaging, every reader reads everything. So inter-reader variability needs to be taken into account. Some assessment of that needs to be done again ahead of time in order to power the study appropriately.

So the focus of intra- and inter-reader variability is a little different when it comes to diagnostic imaging versus imaging in therapeutics.

DR. SRIDHARA: I must say I am stepping outside of my comfort, but I can tell you that we have assumed that each reader reads everything on all patients for oncology. So at least that is our assumption that that is what is

happening.
If the idea is that many readers can read, then you have all these site readers already, and we have already discussed at length that they are not trained properly or there are differences and so on and so forth.
So, if you have half the patients being read by one reader versus the other, then maybe they are not seeing the same things. We don't know that.
The other thing is if you assume they are all experts, they are trained identically, so they are supposed to be reading exactly the same way, then there is no adjudication. That is what we are saying, that we have come to a point where the expertise and the standardization that has happened, that there is no adjudication required at all. That would be a perfect scenario, but I don't know if the technology is there yet to do that.
DR. ZALKIKAR: From a statistical point of view, if you don't have any missing data, have low, low adjudication rate, have very low rates of discrepancy, and we won't have perfect datasets to look at.
DR. STEVENS: But having said that, since we are talking about statistical analysis plans, the thing that is really important to understand here is that all of these contingencies need to be built into your statistical analysis plan; that is, if you anticipate having a reader not be able to read all of the images, that needs to be addressed in the analysis plan as to how you are going to handle that situation, how you are going to change the analysis of the data, how you are going to factor in the readers, how you are going to do all of these things, how you are going to make sure that there is no bias introduced into the system, because one reader has already read a bunch of images, and theoretically, you guys know the results of those images before they are done, and you are not going to retrain the other reader to read them differently.
So all of those issues really need to be addressed up front before you get started. Otherwise, when the reader drops out, you have a post hoc analysis on all of your data, and then all the analysis becomes suspect from that point forward.
DR. ZALKIKAR: Right. I should add prespecify everything.
DR. DORFMAN: Just so that it can end up in the minutes, I didn't see here, but I think as opposed to the IRC needs to be in the statistical analysis plan, how to deal with the issue that was raised yesterday, which is the issue of unscheduled exams.
Yesterday, we talked about it from the standpoint of how identical do they need to be before you allow them to be considered as datapoints, but today, we have been talking about time to progression.
For some types of illnesses, like the rheumatology therapies, we are probably going to have fewer unscheduled exams, but certainly, for neurologic diseases, cardiovascular diseases, and most likely for oncology diseases, there may be unscheduled exams that are more common, and because of either cost, contrast or radiation and the fact that they fall close by, they may replace, if you will, on protocol and scheduled exams.
Clearly, somewhere that area of the accrual and data points which will be in one part of the protocol will need to be dealt with in the statistical analysis plan, and it would seem to me that the charter is a good place at least to make sure that the two of them square off, one against the other, even if it is only by referring to them and that by

doing so, someone will review the fact that they are both squared off.
Statistically, again, I am not sure how. I know there are lots of ways to deal with it, but I would assume that would be explicitly stated in advance because this will be a big problem in most of these trials.
DR. RAUNIG: I would like to address the rheumatology issue here of volume measurements or thickness measurements that don't necessarily have to do with the diagnostic measurement.
The important part of the intra-reader variability and the volume measurement or why you might want to have -- there is a drop-off or a cut-off point for making the decision of whether to have the same reader read the same patient or not. That is a correlation of 50 percent. That is the cut-off point. So anything below 50 percent correlation, it doesn't matter whether you have the same reader read the same patient or not. Anything above 50 percent, this is the intra-patient correlation within a reader. Then you should have the same patient being read by the same reader, and 50 percent is the cut-off. If you get to a correlation of .7, then you really, really should have the same reader read the same person, and that correlation really has to do with the intra- and inter-reader variability.
The correlation is simply the radio of the inter-reader variability to the total variability or the sum of both of those. So that is one way that you can actually determine, or that is how the statisticians determine whether we would look at a changed score determined by a single reader or whether we don't need a single reader at all and we can randomize all reads or all images over all time points and just have multiple readers read multiple time points.
MR. GASTINEAU: We will take one more question and then try and break for lunch.
DR. NOURSALEHI: I was just going to add that in situations where inter-reader variability is a concern, I don't think it is sufficient to have the same reader read a complete patient. That is not sufficient where the inter-reader variability is a concern because whether that patient is in an active arm versus a placebo arm, that makes a big difference. So the consistency has to be applied across all treatment arms.
What we have done in our studies is to ensure that the same reader at each site reads all the patients, regardless of which arm they are on.
In cases where you have thousands and thousands of images, it may be really not practical for the same reader to read everything, but at least we can ensure that happens in the same site.
MR. GASTINEAU: Okay. David?
DR. RAUNIG: I would like to address that very quickly.
The only problem you have with that is you completely confound reader by site, which actually would be okay. You would increase the variability of the site because that is subsumed into there.
Just a slight modification of what you said, but I know what you are saying, it is not the inter-reader variability that matters. It is the intra-reader variability. Once the intra-reader variability becomes a larger and larger percentage of the total variability, that is when you don't really need the same reader reading the same patient.
DR. NOURSALEHI: If I may add, the intra-reader variability is training and consistency, that throughout the

study you have to make sure there is no drift in the read. So that is a different issue, but inter-reader variability, you can address that by making sure that the same reader has a balance between the arms of the study.
So, if you look at the active arm, the same reader has to look at the placebo arm. So you assess both sides of the spectrum by the same standards by the same reader. That is your inter-reader variability.
The intra is purely by training and staying on top of issues with all the readers.
MR. GASTINEAU: Okay. Thank you very much to our panel. Thanks for all the interactions.
We will take a break for lunch and come back this afternoon.
DR. MOZLEY: Good afternoon again, everyone. Thanks for coming back.
You may have noticed that we are running this program as a continuous process of adaptation that has been designed to foster communication and allow everyone equal opportunity to contribute to the meeting and take from it what you will.
At this point, I would like to ask the group's permission to please consider revising the schedule one more time. We already made a group decision to combine the breakout sessions and stay as one herd. At this point in our meeting's life cycle, I would like you to think about allowing the site interface group about 15 minutes to finish their presentation and bring forward some points that they think are quite meaningful.
The proposal would then be to bring the FDA up for an informal panel discussion. We would be very interested in hearing what they thought the meeting might have accomplished and what was still undone. It would be of some interest to hear where they thought we came in with some misalignments, where some of those alignments might have improved, and what work still needs to be done, and we are going to do is in a very informal collegial way, to the extent we can.
We would then like to bring the session subgroup co-chairs back up to help the whole group formulate an action plan, that is, to create a path forward that we all agree to.

We need to explicitly specify what some of our deliverables will be, how we will construct those deliverables, and what some of the timelines might be.
Then it being about 3 o'clock, maybe a little after, we would adjourn and go our separate ways.
So that is the proposal, and I am asking your permission to at least consider it as the default. It boils down to an accelerated afternoon, a more interactive afternoon, an afternoon that is more focused on the future than the content of the current white papers, which we have all read many, many times. You can make a data-driven decision as to how long we stay, as long as it is useful. So would that be all right?
We have spoken about this, a lot of us, one on one, and there does seem to me to be a strong consensus that this plan is favored, but I want to make sure that nobody feels like they are being coerced into accepting something that will ultimately lead to an early adjournment if they don't feel that is in their interest.
So we will see how it goes. Let's please start by asking Dr. Clunie and Dr. Ashton, et al., their group to please

complete their presentation.

Session 3A

Recap and Next Steps: Key Points to Consider for Core Laboratories

DR. CLUNIE: So Ed will be the next speaker. We will just quickly scroll to his slides.

It is all yours.

DR. ASHTON: I am going to be brief. I have only got a few slides here, and I know it is possible to spend two hours on one slide, but we are not going to do that here.

There are just a few points that we worked into the document, looking at how you ought to in the consensus view of the group look at getting sites initiated, getting sites ready to run, making sure that they can support a particular protocol, and then during the course of a study, making sure that they are continuing to effectively support the protocol.

These first two points really involve communication with the radiology folks, making sure that you have a communication plan in place, that the radiology people at your sites are aware that there is a trial ongoing, that that information is not restricted just to the PIs.

Again, I think these are all points that most people in the room probably agree on ought to be done, and maybe they are not being actually implemented particularly well in some trials.

Beyond that, it is, of course, very important -- again, I don't think there is any real disagreement about this -- to make sure that training is provided to site radiologists which is appropriate, to make sure that they are able to support this study. How that training occurs is going to be very therapeutic area-dependent and very dependent on the particular protocol that you are implementing.

If you are running a protocol which is essentially site standard of care, probably the training is going to be limited primarily to things like deidentification, some of the issues that we covered earlier this morning, and how to get the data transferred appropriately.

If you are running a more complicated protocol which differs substantially from what the radiologists typically do as their standard of care, you may need to go so far as to have people on site to do training.

Of course, as everyone has mentioned, we need to make sure that we have an operations guide, but maybe more importantly, make sure that that operations guide is actually read and understood by the folks at the site, and that, again, is a piece that I think might be missing in some trials as they are carried out right now.

So just having the document there, particularly if it is kind of an obtuse legalese sort of document maybe is not sufficient to make sure the site is going to be able to carry out the trial effectively, and of course, site compliance needs to be evaluated at some point, whether it is through phantom studies, test transfers, or just looking at the first few patients that a particular site produces to make sure that they are evaluating things properly.

This last point, specifying in the IRC whether protocol violations at baseline should be duplicated at follow-up, that may sound a little strange, but the issue that we are addressing here is the question of if a baseline scan is not imaged properly and, for various reasons, can't be reimaged, what do you do with the follow-ups. Again, I think how this is handled is going to vary by study.

If you look at a resist study and you were intending to get a chest and a pelvis and you only get the chest and abdomen at the baseline, I think it is pretty clear that you just go ahead and get the chest and pelvis at the follow-up, and you can't evaluate pelvic lesions probably as target lesions in that case.

However, if you look at a more complicated sort of study looking at blood flow or something else and the incorrect protocol is used at baseline, in order to get comparable data, you may have to use that incorrect protocol for that patient throughout the study. The only point here is that that really needs to be specified.

Now falling into the category of being unable to resist picking at a sore, I did want to talk briefly about phantom evaluations.

Again, there was some discussion about this yesterday, and there is some disagreement I think as to the value and the necessity of running phantoms. Certainly, again, this is going to be study and therapeutic area-specific. However, there certainly are some circumstances where it is clearly not appropriate to run phantoms.

Again, if you are doing a CT resist study, you are probably wasting money if you are doing phantom studies in that case. However, there are other cases where it probably is appropriate, and this is one example.

So, for instance, if you are looking at things like cortical thickness or joint space or other things with MRI, this sort of phantom you might want to run, and these are actual phantom runs that we have gotten at site initiation meetings from various sites.

This is the uniformity and linearity phantom, and on the left is what it is supposed to look like, and the right is an actual scan that we pulled from one of our site initiation visits.

As you can see, if you are trying to precisely quantify things using an MRI scanner, this scanner is probably not going to do the job for you. This sort of addresses the question, again, that was raised yesterday that sites have their own QA standards, so why do we need to pile on top of that. The answer is that in many cases the QA standards that they have at the sites are really not adequate to what we need to do.

They may be perfectly adequate if you are just doing evaluations, subjective interpretation, but if you are trying to quantify things, which we do in some of our trials and we don't in others, you may need to assess whether the scanners are capable of providing that level of quantification.

Another area where this becomes an issues is when you are doing things like dynamic contrast, enhanced MRI, other things where you are looking at relating changes in signal to actual physiological changes.

So, again, this is an example from one of our DC MRI studies where we are trying to look at changes in gadolinium concentration within the body and relating that to changes that we observed in the signal in these patients, and what we like to see is what we see on the left, which is a really nice linear relationship between the two with no real weighting to the initial T1 of the issue that we are examining, that the scan on the right, the phantom, you can see not only do we not have this nice linear relationship, in fact, we have a non-monotonic relationship and a huge dependency on the initial T1. The reason for this, this is actually the same protocol, just

using different coils.

So, again, you need to make sure that the equipment that the site has -- and the reason for this, incidentally, was because this is an HPA coil which had two of the elements were failing, but the site didn't mind this, and it still was producing images that were good enough for their evaluations, but if you are trying to do, again, quantification, that may not be, in fact, good enough.

PET is another area. It is an area where I think a lot of people have the impression that it is a little simpler than MRI, but you actually do have some of the same issues, and I want to give thanks to Helen Young from AstraZeneca who gave us permission to use this particular data.

Again, if you are looking at PET and quantifying SUV, you want to make sure that the SUV you are getting from a particular scanner is going to be comparable to the SUVs you are getting from other scanners, and that is actually reflects reality. Of course, there are a number of phantoms that are widely available that you can use to do this if you take the care to do it, and this is, again, an example from -- I blanked out all the site information here, but this particular site study, we calculated.

If you look at the last box on the top grid, which probably none of you can actually read, but I can tell you it is about 34,000 per milliliter, that is the known concentration of activity in this particular phantom at the time this scan was required, but then when we calculate based on the reconstructed data, the concentration, we get something on the order of about 27,000 in this case. So we are off by about 18 percent in this case, which is really completely unacceptable. We like this to be within, certainly, 10 percent of the nominal concentration, and again, this is a scanner that was passing the site QA, but if you calculate an SUV using this scanner, you are going to get something that is 20 percent lower than it really ought to be.

And not only that, because we have been doing this for a while and we have looked at a very wide variety of sites and we look at these sites over time, you see drifts in these numbers. So, if you have the result of the scanner drifting and, for instance, trending lower, you are seeing results which are spurious to what is actually going on biologically with your patients. So, in cases like this, it is important to have these sorts of evaluations.

Once you have done this, once you have effectively qualified a site, you are assured they are doing things correctly, there are some other things that probably need to be done during the course of a trial to make sure that they are continuing to effectively support your trial.

In particular -- and this is a point that was argued a little bit amongst the group, but I think we were able to come to some consensus -- reviewing data for quality purposes as it is collected, in other words, not collected in CDs, dumping them in a bin and then pulling them all out eight months later to do the reads, is probably the right way to go.

You don't necessarily have to do the analysis as the data is collected, but to at least have a quality review. So if a particular site, for instance, is not supporting the protocol appropriately, you catch it after they do it once, not after they have scanned 15 patients and then you have 15 patients worth of data that is lost.

It is important to provide immediate feedback to the sites regarding protocol deviation. So, again, everyone makes mistakes. You would prefer that they only make the same mistake once and not over and over and over before it is noticed and corrected.

Of course, if you notice that sites are making particular mistakes over and over again, it is important to provide additional training, and that also may be important, for instance, in the case where the folks that you trained initially six months ago have now moved on and there is some new person stepping in who doesn't really know what is going on with your study. You may need to do repeated training in that case or in the case where you just see repeated deviations from a particular site, and then it is very important to keep that line of communication which you established at the initiation open throughout the study.

This is particularly important for the case that several people have mentioned where you may have 250 sites and maybe each one is only occurring in a couple of patients and this is over a two-year study. You may go six months in between accruing patients at one particular site. They may have forgotten that your study is even ongoing. So it is important to maintain communication with these folks on some periodic basis and to maintain control over their systems. Again, if that means some periodic rescanning of phantoms to make sure that they still remember how to do the protocol, that will cost a little bit of money, yes, but in many cases, that may wind up being a wise investment.

So that is really all I had to say. I will be happy to take any questions if anyone has any, unless anyone just wants to get straight to the airport, which is fine, too.

DR. ASHTON: Okay. Thank you very much.

DR. CLUNIE: Okay. Aldo has some comments to make, but he is going to reserve them for when the FDA is starting off its panel session. He has just a couple of slides to show.

I am going to hammer through the remaining topics and just give a cursory impression of what is in the document with respect to some. If we have time at the end in the remaining seven and a half minutes, we may be able to fit some comments in.

As you know, I can talk much faster than this. Transfer to the CRO, basically it is in the too hard basket. It is probably going to get solved by the technological solutions that are becoming available and being deployed anyway for other reasons. So we decided basically we were going to focus on deidentification aspects, which will remain a problem until we reach consensus on that.

Compression. Basically, there are no real standards for how much compression is acceptable for any particular task, and there is tremendous variation across tasks.

The scientific literature is riddled with rubbish. It mystifies why the journals publish this stuff, but basically poorly controlled studies, underpowered studies, studies that are based on subject quality assessment instead of observer performance, studies using proprietary one-off algorithms developed by a postdoc that never reached fruition, and a few studies of standard algorithms.

So, basically, it is expensive to do real tests on whether or not lossy image compression matters in a general

sense, and they are very difficult to generalize even when one has conducted them. So people don't.

Our conclusion essentially in the literature review was that there is a strong suspicious that a moderate amount of lossy compression, where a human is performing interpretation task, probably doesn't matter, but there is increasing evidence that any lossy compression affects the performance of machine-driven algorithms because most compression schemes are tuned to the human visual system, if for no other reason. So our recommendation is don't.
If you are going to use it in a trial, our recommendation was to provide statistically sound evidence in the IRC that it was okay to use it, and nobody is going to be able to reach that bar, so it is not going to happen.

DR. CLUNIE: If lossy image compression is inherent in the acquisition process, if, for example, you are on an echocardiography machine which spits out video signals that are compressed as they come off the chip, then you have no compressed or uncompressed predecessor. That is a different case, and that is the standard of care. So we have to accept that, but that is not what CT scanners do. That is not what MRI scanners do. Otherwise, don't use it.
If the site wants to use it, tell them not to. If the PACS lossy compresses automatically before the person has control over it, get the images directly from the modality and take the PACS out of the loop. Every modality has either a network connection, a CD burner, or an MOD burner, and with discipline and early intervention, you can get the images. You just need to train the sites.
Also, make sure that your QC process, you are checking the necessary DICOM header flags to see if there is indication that lossy compression has been used because it may be hard to tell simply by visual inspection.
There were a number of comments made on our compression recommendations. Essentially, when the FDA made a number of comments, they were really questioning the fact that the draft asserts that specific trials require specific evidence, which implies that the burden is on the sponsor or the CRO to provide this, and they questioned that, and that is indeed what we intended the draft to say. So our current consensus is that compression is not yet cool, and it should be avoided in the context of these kind of trials until somebody proves otherwise, regardless of whether or not it becomes common practice in clinical care.
Digitization, I won't dwell on. Basically, digitization is to be avoided if you have digital material to start with. If you are faced with film screen x-ray modalities where they are not using CR, they are not using direct digital detectors, then you really have no choice. The original material in this case is film. It has to be digitized. There are ACR guidelines on the subject which may or may not be worth the paper they are printed on.
Clearly, it is important to use medical-grade equipment here. You just don't buy a $100 flatbed USB scanner and put it on your desk and scan x-rays here, and the FDA specifically raised the question do you actually need to

use an approved medical device for this purpose. I think it is fair to say that all x-ray scanners are approved medical devices. So the question is somewhat moot, but it is a good point to fall back on.

Which brings up the question of equipment in a general sense, not something we have yet addressed in the document, and that is the question of do we need to use approved equipment, do we need to use validated equipment.

There is a presumption that the equipment at the site is both reliable, all issues of phantoms aside, and is an approved medical device, which is generally the case, but not necessarily. The site may use a workstation to transfer the images. The site may use a deidentification tool as some freebie they downloaded from the Internet. So a certain amount of caution is required here, and the FDA has made it clear that computerized systems used in clinical investigations need to be validated.

21 CFR Part 11 is often invoked, but there are many subtleties to 21 CFR Part 11 with respect to what requirements are present for audit trails, for signatures, but inevitably, any software that you use in the context of 21 CFR Part 11 or in the context of a computerized system for clinical investigation, particularly one that results in registration, is the need to validate it.

This becomes complex when you are trying to instrument the sites with software or influence them to choose certain software, and it is also complicated in a multi-center, particularly international environment where for the purposes of audit trails and signatures, you need the identify of the person performing the work.

So it is very easy for us to stand up here and say when they do deidentification at the site, they must create an audit trail, but that audit trail implies their identify. How do we track their identity? How do we authenticate that they are who they say they are? Which makes paper seem very attractive, as we talked about earlier.

A site's own tools in general are likely to be invalidated, unreliable, and undocuments, both in terms of the equipment, the software, and the people who use them, and the people who use them are likely not trained to use them.

Reuse. I will just very briefly touch on reuse. I think everybody knows why we want to reuse images in a context outside of clinical trials, because otherwise the data is essentially wasted, particularly in those trials that fail, and there is both a common good and a self-interest at stake here in the sense that tools need to be improved, tools need data to be improved, and this is a potential source of data.

The mechanics of it really are oriented around minimizing risk, why should a sponsor take on the risk of exposing their data to a greater audience. There are reuse factors internally within a sponsor, of course, but we are really talking about reuse in a broader context, if not a public context, a contribution to government archives, that kind of stuff.

There is the matter of consent, whether it is or is not needed, making sure that the consent is irrevocable if this

data is released into the public domain, the primary importance of avoiding delay in IRB approval, because you have created some kind of funky consent that includes a reuse component, and then there is the question of how it is distributed and a thorough, very thorough deidentification that needs to take place, and then the cost of the QC to perform such more thorough deidentification than might otherwise be necessary.

So the document talks to some of these issues, primarily with a focus on trying to achieve a standardized component to include in a consent form that we could all agree to and IRBs could get used to approving, but a lot of work needs to be done there.

The other discussion point that the FDA kindly mentioned, particularly in reference to the site interface group, is how poorly the material is organized in the existing document, and we take that at face value. It is indeed organized poorly, and there is a lot of editorial work needed in the document, particularly within our set of material, but we also not just need to achieve coherency and consistency with the text within our contribution and between contributions, but also factor out the lexicon from a subgroup, so that it is central and consistent.

And that's it for us. You have 13 seconds for questions.

DR. CLUNIE: I don't think we have time for any questions, unfortunately, but feel free to join the group and participate in the next phase.

DR. MOZLEY: Dr. Clunie, I think it is okay to entertain questions if people really have something meaningful that they think needs to be brought forth to the group. I would just encourage people to keep it abbreviated and essential, if that is all right with you

DR. CLUNIE: Okay. That's good. Well, before I dismiss the panel, are there any brief and meaningful comments?

DR. CLUNIE: Anybody? Yes, there is one.

ATTENDEE: Do we want digital video to fall into the compression and digitization?

DR. CLUNIE: I think the digital video is a prime example of what I was referring to as already compressed or already ruined, depending on how you look at it.

For example, if you are using a DV camera, then the DV machine automatically compresses it. You get it in a compressed form.

If you then perform further compression on it or format conversion, then you raise a very interesting question, are you further degrading it beyond what is necessary, but if it is already compressed -- that is the same question as digitizing film, in essence. If it comes on an analog video chain and you need to distribute it digitally, then you have to perform video capture, in essence, and that is inherently a lossy process by definition and typically generates a lossy compressed format on top of that, but I don't see anything wrong with that. That seems to be the standard of practice both in the trial industry and in the clinical world.

ATTENDEE: Right. There are guidelines from ACR on flat film. There doesn't seem to be anything for motion.
DR. CLUNIE: Right. That is because radiologists don't recognize the existence of motion.
DR. CLUNIE: Yes, I take your point.
Anything else?
DR. CLUNIE: Okay. Well, we are done. Thank you very much.
DR. MOZLEY: You guys are great.
Ted, does your group need a few more minutes to wrap things up, or are you going to be prepared to work with us in the final session on the path forward? You might want to canvas your people real quickly.
Let's ask the audience. Do they need the stats group to come back up and get issues on the record? Remember we are going to put these things on the record, so that we can reflect on them when we go home, and we can broadcast them to people who weren't able to attend in person. So is there something that needs to go on the record for the stats group at this time?
DR. MOZLEY: You still get one more chance. If silence means what I think it does, then I would like to please ask the members of the FDA group who participated in this conference.
Dr. Perrone?
DR. PERRONE: Just for the record, I thought one of the possible next steps of the SAP was to discuss adjudication and thinking about that in a new paradigm, potentially incorporating that discussion in the SAP plan that comes out through the pharmaceutical industry. So if we could just put that on the path forward, I think that might be useful.
DR. MOZLEY: Okay, great. I do think that is an action item for the future, and we would like to capture it in the last session, among others.
I personally can't overstate how appreciative I am of the FDA working with us all these many months. We do understand how busy you are. We do realize that there is a lot of competition for your time. We understand that you sometimes are working under pressure, that a number of constituencies out there, shall we say, are interested in your efforts.
So, again, personally and I think on behalf of the group, we really, really can't thank you enough for coming.
DR. MOZLEY: If you all would please come forward and give us one more chance to have a very informal conversation with you about where we have come from, where we are, where you hope we are going, and how you expect us to get there, we will appreciate that.